

(19)



(11)

EP 3 568 778 B1

(12)

EUROPEAN PATENT SPECIFICATION

(45) Date of publication and mention of the grant of the patent:

10.03.2021 Bulletin 2021/10

(21) Application number: **18741064.2**

(22) Date of filing: **16.01.2018**

(51) Int Cl.:

G06K 9/00 (2006.01)	H04L 29/08 (2006.01)
H04L 29/06 (2006.01)	G06N 5/02 (2006.01)
G06N 3/08 (2006.01)	G06K 9/62 (2006.01)
G06F 3/01 (2006.01)	G06F 11/30 (2006.01)
G06F 11/34 (2006.01)	G09G 5/14 (2006.01)

(86) International application number:

PCT/KR2018/000738

(87) International publication number:

WO 2018/135833 (26.07.2018 Gazette 2018/30)

(54) **SYSTEM AND METHOD FOR CONTEXTUAL DRIVEN INTELLIGENCE**

SYSTEM UND VERFAHREN FÜR KONTEXTGESTEUERTE INTELLIGENZ

SYSTÈME ET PROCÉDÉ D'INTELLIGENCE CONTEXTUELLE

(84) Designated Contracting States:

**AL AT BE BG CH CY CZ DE DK EE ES FI FR GB
GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO
PL PT RO RS SE SI SK SM TR**

(30) Priority: **19.01.2017 US 201762448325 P**

19.01.2017 US 201762448339 P

16.03.2017 US 201762472497 P

13.12.2017 US 201715841157

(43) Date of publication of application:

20.11.2019 Bulletin 2019/47

(73) Proprietor: **Samsung Electronics Co., Ltd.**

Suwon-si, Gyeonggi-do 16677 (KR)

(72) Inventors:

- **ANTOL, Stanislaw**
Gyeonggi-do, 16677 (KR)
- **BENDALE, Abhijit**
Gyeonggi-do, 16677 (KR)
- **GIBBS, Simon J.**
Gyeonggi-do, 16677 (KR)
- **JEON, Won J.**
Gyeonggi-do, 16677 (KR)

- **KANG, Hyun Jae**
Gyeonggi-do, 16677 (KR)
- **KIM, Jihee**
Gyeonggi-do, 16677 (KR)
- **LI, Bo**
Gyeonggi-do, 16677 (KR)
- **LIOT, Anthony S.**
Gyeonggi-do, 16677 (KR)
- **LUO, Lu**
Gyeonggi-do, 16677 (KR)
- **MISTRY, Pranav K.**
Gyeonggi-do, 16677 (KR)
- **YING, Zhihan**
Gyeonggi-do, 16677 (KR)

(74) Representative: **Gulde & Partner**

**Patent- und Rechtsanwaltskanzlei mbB
Wallstraße 58/59
10179 Berlin (DE)**

(56) References cited:

US-A1- 2008 147 730	US-A1- 2014 059 458
US-A1- 2014 253 588	US-A1- 2014 253 588
US-A1- 2015 213 058	US-A1- 2015 347 488
US-A1- 2016 063 692	US-A1- 2016 117 862

Note: Within nine months of the publication of the mention of the grant of the European patent in the European Patent Bulletin, any person may give notice to the European Patent Office of opposition to that patent, in accordance with the Implementing Regulations. Notice of opposition shall not be deemed to have been filed until the opposition fee has been paid. (Art. 99(1) European Patent Convention).

Description

Technical Field

[0001] One or more embodiments relate generally to artificial intelligence and machine learning, and in particular, to use of contextual information to enhance computer vision intelligence, such as selecting appropriate models and recommending actions.

Background Art

[0002] Electronic devices host a set of physical sensors and software mechanisms that have the ability to detect or estimate many elements of context. Cloud based services can further enrich the context.

[0003] There is a research effort to provide intelligent services utilizing such physical sensors, software mechanisms and services.

Disclosure of Invention

Brief Description of Drawings

[0004]

FIG. 1 shows an electronic system for contextual driven intelligence, according to some embodiments;

FIG. 2 shows an example architecture for contextual driven intelligence, according to some embodiments;

FIG. 3 shows an example diagram for context-based model selection, according to an embodiment;

FIG. 4 shows an example use for context-based application launch, according to an embodiment;

FIG. 5 shows another example use for context-based application launch, according to an embodiment;

FIG. 6 shows an example of context amplification including a context amplification feedback loop, according to an embodiment;

FIG. 7 shows example uses of context amplification, according to an embodiment;

FIG. 8 shows an example use for scene recognition and sensor inputs resulting in context detection and application recommendation, according to an embodiment;

FIGS. 9A-C show examples of a user interface for layer activation, according to an embodiment;

FIGS. 10A-D show examples of a user interface for context-based model selection and activation with different layers, according to an embodiment;

FIG. 11 shows a block diagram for a process for contextual driven intelligence, according to some embodiments;

FIG. 12 is an exemplary high-level block diagram showing an information processing system comprising a computer system useful for implementing dis-

closed embodiments; and

FIG. 13 shows a block diagram for a process for context-based model selection and activation, according to some embodiments.

Best Mode for Carrying out the Invention

[0005] The present invention is set out in the appended independent claims. Preferred embodiments are set out in the dependent claims.

Mode for the Invention

[0006] The following description is made for the purpose of illustrating the general principles of one or more embodiments and is not meant to limit the inventive concepts claimed herein. Further, particular features described herein can be used in combination with other described features in each of the various possible combinations and permutations. Unless otherwise specifically defined herein, all terms are to be given their broadest possible interpretation including meanings implied from the specification as well as meanings understood by those skilled in the art and/or as defined in dictionaries, treatises, etc.

[0007] One or more embodiments provide for retrieving, by a device, contextual information based on at least one of an image, the device, user context, or a combination thereof. At least one model (where the term model may refer to a machine learning model, deep learning model, recognition model, etc.) is identified from multiple models based on the contextual information and at least one object recognized in an image based on at least one model. At least one icon is displayed at the device. The at least one icon being associated with at least one of an application, a service, or a combination thereof providing additional information. It should be noted that the terms "at least one of" refers to one or more than one of the elements that follow. For example, "at least one of a, b, c, or a combination thereof" may be interpreted as "a," "b," or "c" individually; or as "a" and "b" together in combination, as "b" and "c" together in combination; as "a" and "c" together in combination; or as "a," "b" and "c" together in combination.

[0008] FIG. 1 shows an electronic system 50 that may be employed with one or more embodiments for contextual driven intelligence. The electronic system 50 includes a first device 52, such as a client or a server, connected to a second device 56, such as a client or server. The first device 52 may communicate with the second device 56 with a communication path 54, such as a wireless or wired network.

[0009] In some examples, the first device 52 may be of any of a variety of networked devices. The first device 52 may couple directly or indirectly to the communication path 54 to communicate with the second device 56 or may be a stand-alone device.

[0010] For illustrative purposes, the electronic system

50 is described with the first device 52 as a display device, although it is understood that the first device 52 may be a variety of different types of devices. For example, the first device 52 may also be a device for presenting images or a multi-media presentation. A multi-media presentation may be a presentation including sound, a sequence of streaming images or a video feed, or a combination thereof. As an example, the first device 52 may be a smart telephone device, wearable device, tablet device, camera device, video device, portable computing device, kiosk, gaming device, ultra-high definition TV (UHDTV) device, or any other type of display device (e.g., monitor, video panel, heads up display (HUD), etc.).

[0011] The second device 56 may be any of a variety of centralized or decentralized computing devices, image or video transmission devices. For example, the second device 56 may be a smart TV, a multimedia computer, a tablet, a laptop computer, a desktop computer, a video game console, grid-computing resources, a virtualized computer resource, cloud computing resource, routers, switches, peer-to-peer distributed computing devices, a media playback device, a Digital Video Disk (DVD) player, a three-dimension enabled DVD player, a recording device, such as a camera or video camera, or a combination thereof. In another example, the second device 56 may be a signal receiver for receiving broadcast or live stream signals, such as a television receiver, a cable box, a satellite dish receiver, or a web enabled device.

[0012] The second device 56 may be centralized in a single room, distributed across different rooms, distributed across different geographical locations, embedded within a telecommunications network, etc. The second device 56 may have a means for coupling with the communication path 54 to communicate with the first device 52.

[0013] For illustrative purposes, the electronic system 50 is described with the second device 56 as a computing device, although it is understood that the second device 56 may be different types of devices. Also for illustrative purposes, the electronic system 50 is shown with the second device 56 and the first device 52 as end points of the communication path 54, although it is understood that the electronic system 50 may have a different partition between the first device 52, the second device 56, and the communication path 54. For example, the first device 52, the second device 56, or a combination thereof may also function as part of the communication path 54.

[0014] The communication path 54 may be a variety of networks. For example, the communication path 54 may include wireless communication, wired communication, optical, ultrasonic, or the combination thereof. Satellite communication, cellular communication, BLUETOOTH®, Infrared Data Association standard (IrDA), wireless fidelity (WiFi), and worldwide interoperability for microwave access (WiMAX) are examples of wireless communication that may be included in the communication path 54. Ethernet, digital subscriber line (DSL), fiber to the home (FTTH), high-definition multimedia interface

(HDMI) cable, and plain old telephone service (POTS) are examples of wired communication that may be included in the communication path 54.

[0015] Further, the communication path 54 may traverse a number of network topologies and distances. For example, the communication path 54 may include direct connection, personal area network (PAN), local area network (LAN), metropolitan area network (MAN), wide area network (WAN), a combination thereof, etc.

[0016] One or more embodiments may provide processing for context-based model selection and activation of corresponding actions; context-based application launch by "real-world browsing" that directly takes a user from a camera view to an application or a web view associated with the recognized object; context amplification by activating additional models and new sources for visual context, which may in turn lead to further model activation; and additional sources of context by identifying user state and predicting user intent from multiple context sources.

[0017] FIG. 2 shows an example architecture 200 for contextual driven intelligence, according to some embodiments. With advances in computer vision technology and increased computational power of electronic devices, the camera 205 itself is becoming a context sensor. The camera 205 (e.g., an intelligent camera) detects objects and activities in the vicinity of a user. For example, camera imagery may be used to classify the location of the user ("home," "office," "beach," etc.), to detect common objects, and to recognize specific faces and emotions. The intelligent camera 205 derives from advances in computer vision and artificial intelligence that make use of deep learning technology. Deep learning refers to a set of computational techniques for training deep multi-layer neural networks on very large datasets. These trained networks ("models") may then be run on low-cost hardware such as that found in mobile phones. Contextual information may be used to enhance computer vision intelligence, such as selecting appropriate models, recommend actions that are relevant to the user, etc. As described below, reference is made to models, "DL models" or deep learning models, however, these models are referred to for just some embodiments and other types of models may also be used, such as those based on traditional machine learning or computer vision techniques. In one embodiment, the architecture 200 may use a combination of DL models and traditional models.

[0018] In one embodiment, architecture 200 includes an electronic device (e.g., a mobile phone device, tablet, wearable device, a smart camera device, mobile computing device, etc.) including a camera 205, DL processing element 210, a context generation processing 220, intent prediction processing 240, layers 250, layer activation and recommendation processing 255, camera application 260, sensors 230, speech/voice input 231, Internet or network connection 232, and installed applications 233.

[0019] In one embodiment, the DL processing element

210 may include visual classifier results 211, which are outputs from one or more deep learning engines and corresponding models. Model management processing 212 is responsible for monitoring and selecting/deselecting multiple models (e.g., model 1 - N, N being a positive integer), as well as maintaining the lifecycles of each model, including initialization and termination. In one embodiment, the visual classifier results 211 are input to the recommendation processing 255. The context generation processing 220 may include a context analyzer 221 that receives context from static context 222, dynamic context 223, learned context 224, and inferred context 225. The static context 222 receives contextual information from static sources 226, for example, static user profile information, user preferences, device settings, etc. The dynamic context 223 receives contextual information 227, for example, global positioning system (GPS) information, time of day, user activity, weather, web events, etc. The learned context 224 receives contextual information 228, for example, application/app usage, application/app activity, device interaction, etc. The inferred context 225 receives contextual information 229 such as health activity, application/app usage patterns, browser cache, cookies, etc.

[0020] In one embodiment, the sensors 230, speech/voice input 231, Internet 232, and installed applications 233 may reside within a same electronic device as the camera 205, or be distributed among multiple electronic devices.

[0021] FIG. 3 shows an example diagram 300 for context-based model selection, according to an embodiment. In one embodiment, an activation engine 310 (of the layer activation and recommendation processing 255, FIG. 2) manages a set of layers/models 330, and receives information from context generation processing 220. Each layer (from layers 250, FIG. 2) comprises a provider (e.g., a mobile application, a web service, or a collection of information such as a database), a visualization 331 (e.g., an icon representing the provider), associated layer activation rules 335 and a model 336 (e.g., ML model, DL model, recognition model, etc.). In one embodiment, layer activation has two parts: model selection: determining when the layer's model 336 (if present) should be started (which means it will receive visual data, such as frames from the camera 205 (FIG. 2) or other sources of imagery); and provider selection: determining when the provider and visualization 331 of the layer should be presented. Once a layer is activated, the layer visualization 331 appears and the user may select the layer. This selection may be via touch (e.g., the user taps on the layer icon on a touchscreen display of an electronic device, (e.g., a smartphone), voice (e.g., the user speaks the layer name) or by other forms of interaction such as pointing (e.g., via a mouse, a pen device, etc.). When the layer is selected, the provider presents information related to the context element (such as a recognized object) that led to activation of the layer.

[0022] When large numbers of layers (with associated

models) are placed on an electronic device (e.g., device 52, FIG. 1), not all layers can be active at the same time since device resources (screen space, memory, central processor unit (CPU) and power) are limited. Thus, a key problem is selecting what layers/models to activate. One solution is to let the user select - but this puts a burden on the user. Furthermore, this burden grows as the number of layers increases. In one embodiment, another solution is context-based model selection and layer activation, which allows a system running on the electronic device to recommend layer/model activation based on contextual element. In one embodiment, the activation engine 310 is responsible for managing layer activation. A layer's activation rule is used by the activation engine 310 to determine when the layer should be activated. The activation rule can take many forms such as a rule-based system or a classifier built using traditional machine learning techniques.

[0023] Example diagram 300 illustrates one embodiment of context-based model selection and layer activation. The context may include, but is not limited to, predicted visual labels (e.g., "computer mouse") from the currently running model(s) plus other forms of context such as location 321, intent 322, user preferences 323, audio context 324 and visual context 325. One example of contextual model selection and layer activation may include: visual label and layer keyword similarity above threshold. In this example, layer activation is triggered when a generic model predicts a visual label similar to a keyword associated with the layer. This layer is therefore activated and its associated specific model 336 is selected. For example, a shopping layer with icon "a" might be activated if some visual label is similar to keywords such as "shopping," "electronics," "books," etc. Techniques for implementing semantic similarity between words or word phrases include distributed word vector models (e.g., word2vec) which convert words into (relatively) low dimensional vectors. Word vector models provide a good job grouping word meaning, thus things such as dot product (sum of element-wise multiplication) can capture similarity between words. For example, suppose the word vector for "computer mouse" is [0.1, -0.9, 0.0], and "electronics" is [0.2, -0.8, -0.1], "food" is [-0.5, 0.1, 0.9], then using the vector dot products, "computer mouse" is more similar to "electronics" (0.74) than "food" (-0.14). Note that for illustrative purposes, three element vectors are used, but in practice much longer vectors are used. It should be noted that generic models are started automatically and kept running. Generic models have a layer activation rule just like the non-generic models. "Objects" (general object classification) and "Scenes" (general scene classification) are examples. The generic models reside in the same place as other models (e.g., model management processing 212, FIG. 2) and layers (layers/models 330, FIG. 3). In the user interface of one embodiment, (e.g., FIGS. 9A-C) the layers are ordered such that non-generic layers appear first (at the top of the list).

[0024] Another example of contextual model selection

and layer activation may include: visual label confidence above threshold. In this case, layer activation by the activation engine 310 is triggered when a model predicts a visual label with confidence above a certain threshold. For example, the "Food Network" layer could be activated if the "food model" recognizes something with confidence above 90%.

[0025] Yet another example of contextual model selection and layer activation may include: visual label in specified set. In this example, layer activation by the activation engine 320 is triggered when a model 336 predicts a visual label, with confidence above a certain threshold, and appearing in a specified set. For example, the "National Geographic" layer could be activated if a model predicts a "tiger" label or a "lion" label.

[0026] Still another example of contextual model selection and layer activation may include: matches to other forms of context (e.g., location, intent, user profile). A layer may be activated by the activation engine 320 if other (non-visual) parts of the contextual element match certain values. For example, the "USED CAR SERVICE" layer might be activated if the user is located in a car lot based on GPS data.

[0027] Another example of contextual model selection and layer activation may include: combinations of the above examples. Model selection and layer activation by the activation engine 320 may be determined by some logical combination of the above. For example, when the user is at a grocery store and a grocery product is detected; then a layer providing nutritional information may be activated.

[0028] In one embodiment, third-parties can bid for keywords, location, etc. to obtain higher visibility on the device. Suppose, for example, a car company "F" has obtained the keyword "car" and provided an F layer that is installed on an electronic device. This means that if the contextual layer activation system with the activation engine 320 is running and a car is recognized then an F icon will appear and the user is one click away from the F web site or mobile application.

[0029] In one embodiment, users may customize layer activation. For example, a user may have preferred providers for electronics items, clothing, or restaurant recommendations. One of the challenges in implementing contextual activation is that a process is needed to measure the relevance of a particular layer to the current context. In one embodiment, metrics of semantic similarity between words or word phrases is employed. Some advantages of this approach include: many different information sources (e.g., models 336, context, user intent, etc.) can naturally be described using words; and this approach allows an open vocabulary. Because of the variety of models 336 that may be provided by layers, it is impractical to enforce consistent/limited vocabulary.

[0030] FIG. 4 shows an example use 400 for context-based application launch, according to an embodiment. Example use 400 shows an example of model selection and application activation using the architecture 200. For

the same object (a frog 410), if the user's current context is identified as "outdoors, at San Francisco Zoo" and intent is determined to be "visiting," then the "SF Zoo" model will be automatically selected by the activation engine 310 (FIG. 3), recognizing the object as "Oriental Fire-Bellied Toad" in the informational "card" 423 and enables launching the SF Zoo application using layer icon 424. In one embodiment, the screen 420 of an electronic device (e.g., a mobile telephone) includes an augmented reality (AR) selection 421 that executes architecture 200 (FIG. 2) and also provides other app selections as icons 422, such as Internet search (e.g., GOOGLE®), dictionary search (e.g., WEBSTER'S®), video search (e.g., YOU TUBE®), a map application (e.g., GOOGLE MAPS®), etc.

[0031] In one embodiment, model selection is initiated by the user, such as by tapping on the screen 420 (e.g., a first click) of a camera view captured by camera 205 (FIG. 2), which triggers the contextual layer activation by the activation engine 310. This results in a list of activated layers being shown as icons 422 on the screen. The layers may be ordered based on their relevance to the contextual element, and the most relevant may be expanded to show more detail (e.g., the informational "card" 423 appearing at the bottom of the screen 420).

[0032] In one embodiment, ranking of layers may also take into account whether the layer is from a third-party or a built-in "generic" layer. The ranking of layers can be based on various contextual information such as user's profile and preferences, routine behaviors, time, location, and so on. Layer ranking can also integrate commercial considerations, a third-party layer may rank higher if the layer provider has a marketing contract for product promotion. For example, if the user taps on the layer icon 424 appearing in the informational "card" 423 (e.g., a second click), the provider for the layer is launched. In one embodiment, the provider may be an application on the electronic device or a web site 430 accessed via a browser on the electronic device. If the application or web page supports "deep linking" (for example, exposes uniform resource locators (URLs) with parameters for search results), then the launch may take the user directly to an application view or web page associated with the visual that triggered the launch. This form of interaction is referred to as "real-world browsing" since the user simply taps on an object in the camera 205 view on the screen 420 to bring up an application view or web page (e.g., web site 430 web page) related to the object. In another embodiment, the activation engine 310 runs continually and is constantly fed new frames from the camera 205. In this case, layer icons 422 appear spontaneously and only a single click is needed to launch the provider.

[0033] FIG. 5 shows another example use 500 for context-based application launch, according to an embodiment. In another context, the architecture 200 (FIG. 2) determines using the context generation processing 220 that the user is "indoors, at library" and "reading." The activation engine 310 (FIG. 3) selects the "NatGeo" mod-

el, which recognizes the same type of toad view 410, but shows the "NatGeo" icon 524 that enables launching the National Geographic web page 530.

[0034] FIG. 6 shows an example 600 of context amplification including a context amplification feedback loop 630, according to the claimed invention. In this example, as layers are activated, additional models are started and new sources for the visual context become active. The visual context becomes enriched by these new sources which may in turn lead to further layer activation. Thus, there is a context amplification feedback loop 630 between the context generation processing 220 and layers/models 330 which has the effect of "amplifying" the context. The example 600 shows a processing flow where physical sensing 620 (using sensors 230) and device usage information 625 are input to the context generation processing 220. Here the "Model Lifecycle Manager" 610 of the model management processing 212 (also including the activation engine 310) receives requests to start models 336 of the layers/models 330 from the activation engine 310. The model lifecycle manager 610 is responsible for initialization, and termination of models. The model lifecycle manager 610 may also play a role in load balancing (e.g., throttling recognition requests sent to models). In one example, currently running models 336 of the layers/models 330 identify objects, visual labels are added to the context, additional layers are activated and further models 336 are activated, and more visuals are identified from newly activated models.

[0035] FIG. 7 shows example 700 including exemplary uses 710, 720 and 730 of context amplification, according to the claimed invention. In example use 710, suppose a fairly general object recognition model is running and it detects "food" 712 and "logo" 711 visual labels and the contextual element is "at restaurant" 715. This could lead to activation models 716 including a "food" recognition model specialized for objects found in restaurants, and a "logo" model specialized for logo recognition. When these new models are started in use model 720 (food and logo), it might contribute labels 725 such as "hamburger" and "MCDONALD'S®" to the visual context, which in turn lead to further model activation in use 730 (MCDONALD'S®) such as MCDONALD'S® model 740 being activated. Activation of MCDONALD'S® model 740 brings up the MCDONALD'S® layer containing UI elements such as application icon of MCDONALD'S® 742 in the information card 741. Clicking on icon 742 will open MCDONALD'S® application if it is already installed on device. Otherwise, clicking on icon 742 can either prompt the user to install MCDONALD'S® application, or open MCDONALD'S® web site to provide further information of the recognized Big Mac™. A clear benefit of this approach is that the architecture 200 (FIG. 2) is adaptive. There is no need for a complicated model manager that decides which models to run and which to stop. Instead the architecture 200 automatically makes these choices based on the context.

[0036] In some embodiments, a wide range of sources

to model and infer context information are employed. This information is sourced from multiple sources within the electronic device or from the Internet 232 (FIG. 2). In one embodiment, from the electronic device, there may be raw information and processed information from various sensors 230, voice commands/speech input 231 and installed applications 233. In another embodiment, from the Internet 232, information may arise regarding current activities, weather and other data. In yet another embodiment, various sources may be combined to generate multi-modal context, which is used to predict intent (e.g., using intent prediction processing 240, FIG. 2) and state of the user.

[0037] FIG. 8 shows an example use 800 for scene recognition and sensor inputs resulting in context detection and application recommendation, according to an embodiment. In one embodiment, user intent may be inferred by the context analyzer 221 by identifying the (contextual) state of the user. For example, motion sensors may detect a user's current speed (e.g., walking, running, driving, flying, sitting, etc.), and a GPS sensor may detect rapid change in position of the electronic device (e.g., smart phone), leading to activity detection of the user as "walking." This activity is provided to the context analyzer 221 to generate an intent vector, for example "sightseeing" being the current context of the user. Following this process, the intent prediction processing 240 mechanism provides an actionable signal to the activation engine 310 (FIGS. 3, 6). The results of the visual classifiers are combined with an intent prediction mechanism by a context amplification process (FIG. 6), which is fed to the activation engine 310.

[0038] In one embodiment, the example 800 includes a generic scene recognition result shown on display 810 of an electronic device, that is fed to the context analyzer 221 of the context generation processing 220 (FIG. 2), along with readings from GPS and other sensors 620. On the display 810 is shown the image of the lighthouse along with the information card 423 including icon 811, and icons 422. The resulting display 820 shows the results of detecting that the user is walking near a lighthouse. This piece of location and motion context information is fed to the activation engine 310, which performs semantic analysis and suggests that the user is most likely doing sightseeing, therefore showing icons 830 for recommending relevant travel apps. As shown, the left side of the context analyzer 221 includes: a visual classifier result: light house, learned/ inferred context: outdoors, vacation in near future, and static context: leisure traveler. The right side of the context analyzer 221 includes: intent prediction: trip planning, and recommendations: Travel apps 830.

[0039] In one embodiment, the user interface for model selection and layer activation based upon vision recognition and user context can be used to assist selecting the most relevant contents, filtered from the vast amount of information identified by vision intelligence. In one embodiment, objects detected in the center portion of a

screen may provide more information than the objects in the boundary portions. For example, the information of center objects may include a detailed description while information of a boundary object may only include keywords or indices. In another embodiment, if one object is zoomed in from the screen, AR information will be provided with more details, whilst other objects that decrease in zoom or disappear from the screen have lower priority and little or no information will be provided.

[0040] In another embodiment, objects that gained focus on the screen are provided with detailed information. For example, if it is determined that a user is in a running car, the UI will provide less AR information for outside objects since it is hard for the user to read detailed information with the rapid changes of objects due to the speed of the car.

[0041] FIGS. 9A-C show examples of a user interface when a landmark recognition layer is activated on an electronic device 910 (e.g., a smart phone), according to an embodiment. In one embodiment, in FIG. 9A the screen of the electronic device 910 shows a map 920 and landmark marker 925 on the upper left-hand side of the screen. After identifying the landmark, a set of related layers are displayed on the upper righthand side of the screen, including a messaging app 930. For example, with this recognized landmark, the messaging application can be used to share photos of the location. The information card shows location and icons 422 are displayed on the lower portion of the screen. FIG. 9B shows the screen of the electronic device 910 after activation of the messaging app 930 with a message space 940 for entering a message. FIG. 9C shows a written message 941 as being sent on the screen with a landmark identifier 926 associated with the message 941.

[0042] FIGS. 10A-D show examples of a user interface with various layers activated, according to an embodiment. FIG. 10A shows an object recognition layer (associated with the generic model) activated for an object captured by the camera (e.g., camera 205, FIG. 2) of electronic device 1010. In this example, a controller icon 1020 is selected and a group of apps 1030 are shown for selection. FIG. 10B shows a text and object recognition layer activated for a captured business card 1040, the layer displays a business card recognition app 1021. FIG. 10C shows a location and place (e.g., a landmark, etc.) recognition layer activated for an image 1050 captured by the camera, the layer displays a messaging app 1022. A message 1055 that will be posted at this location is shown on the image 1050. FIG. 10D shows a sign (e.g., a face-and-gesture) recognition layer activated for an image of a girl 1070 making a heart symbol with her hands. The term "heart" is determined based on the girl's hand gesture 1066 and facial expression 1070. This layer displays a set of heart symbol 1065, and a selection of a specific heart symbol 1066 can be used to add to the girl's image 1070. In this example, the selected heart object 1066 may be placed over the girl's hand gesture location to create an augmented effect on the image.

[0043] FIG. 11 shows a block diagram for a process 1100 for contextual driven intelligence, according to some embodiments. In one embodiment, in block 1110 process 1100 provides for retrieving, by a device (e.g., an electronic device 52, FIG. 1, electronic device 910, FIGS. 9A-C, electronic device 1010, FIGS. 10A-D) including architecture 200 (FIG. 2), contextual information based on an image, the device, user context, or a combination thereof. Contextual information includes, but is not limited to: position information obtained using GPS, network-based position, and Wi-Fi data; activity information (e.g., walking, running, sitting, etc.) obtained using motion sensors (e.g. accelerometers, gyroscopes, magnetometers, etc.); audio environment information (e.g. noisy, quiet, etc.) obtained using audio sensors; speech detection and understanding that can detect spoken words and sentences; text analysis that can estimate user sentiment (e.g., angry, happy, etc.), interest, and intent; data analysis that can discover information about the user's preferences and intent (e.g., time, application usage, search history, contacts, etc.) from data collected on device 1010; and cloud services that can further enrich the context (e.g., reverse geocoding that maps from GPS coordinates to addresses). There are many context-aware user intent prediction methods available. In one embodiment, user intent prediction (e.g., using intent prediction processing 240, FIG. 2) can be done by trying to understand the semantic meaning of contextual information, i.e., semantic analysis. One example of semantic analysis that does not require large amount of user data is to use Word Embeddings, where word meanings are represented using vectors and the similarity between two words can be calculated as the distance between the two vectors representing each word. For example, a user context is "walking" results a higher score (shorter distance) with the layer keyword "outdoor" than the layer keyword "restaurant", therefore the "outdoor" layer and its associated models are selected. In block 1120, process 1100 provides for identifying at least one model from many models based on the contextual information and at least one object recognized in an image based on at least one model. In block 1130, process 1000 displays at least one icon at the device. The at least one icon being associated with an application, a service, or a combination thereof providing additional information.

[0044] In one embodiment, for process 1100, the contextual information may include: current time, velocity of the device, sensor information, user speech information, textual information, device application information, and location information. In one embodiment, process 1100 may further include determining at least one learning model relevant to the contextual information.

[0045] In one embodiment, in process 1100 the at least one model is selected from one or more models based on contextual information to provide the additional information (e.g., from an application, a service, or a combination thereof), and the at least one learning model includes one or more additional models (e.g., models 336

(FIG. 3). In one embodiment, in process 1100 displaying the at least one icon at the device is based on at least one of a ranking of relevance of the model, a source of the model, or a combination thereof.

[0046] In one embodiment, process 1100 may further include activating the at least one model based on at least one of the relevance of the at least one model, and receiving input from a user interface of the device. Process 1100 may additionally include displaying a view specific to the at least one model, and enabling an additional model.

[0047] In one embodiment, process 1100 may include determining a visual label based on the contextual information. The contextual information may include camera data. Process 1100 may also include determining the relevance of the at least one model based on the visual label via a computation of at least one of: semantic similarity, visual label confidence, appearance in a specified set, or a combination thereof.

[0048] In one embodiment, process 1100 may also include inferring the contextual information based on data from an external service and sensors connected to the device. The contextual information further includes intent and state of a user of the device. Process 100 may additionally include filtering user-specific content from AR information to obtain the at least one object. In one embodiment, objects in a center portion of a display of the device provide more AR information than objects in a boundary portion of the display, and zoomed in objects provide more AR information than other objects outside a zoomed in portion of the display.

[0049] In one embodiment, process 1100 may include determining the contextual information related to the device being in a moving vehicle, and providing less AR information for objects displayed on the device.

[0050] FIG. 12 is an exemplary high-level block diagram showing an information processing system comprising a computer system 1200 useful for implementing disclosed embodiments. Computer system 1200 may be incorporated in a device 52, 56, FIG. 1, electronic device 910 (FIGS. 9A-C), electronic device 1010 (FIGS. 10A-D), and include architecture 200 (FIG. 2). The computer system 1200 includes one or more processors 1201, and can further include an electronic display device 1202 (for displaying video, graphics, text, and other data), a main memory 1203 (e.g., random access memory (RAM)), storage device 1204 (e.g., hard disk drive), removable storage device 1205 (e.g., removable storage drive, removable memory device, a magnetic tape drive, optical disk drive, computer readable medium having stored therein computer software and/or data), user interface device 1206 (e.g., keyboard, touch screen, keypad, pointing device), and a communication interface 1207 (e.g., modem, a network interface (such as an Ethernet card), a communications port, or a PCMCIA slot and card). The communication interface 1207 allows software and data to be transferred between the computer system and external devices (e.g., over communication

path 54, FIG. 1). The system 1200 further includes a communications infrastructure 1208 (e.g., a communications bus, cross-over bar, or network) to which the aforementioned devices/processors 1201 through 1207 are connected.

[0051] In one embodiment, the main memory 1203, storage device 1204 and removable storage device 1205, each by themselves or in any combination, may store instructions for the embodiments described above that may be executed by the one or more processors 1201.

[0052] Information transferred via communications interface 1207 may be in the form of signals such as electronic, electromagnetic, optical, or other signals capable of being received by communications interface 1207, via a communication link that carries signals and may be implemented using wire or cable, fiber optics, a phone line, a cellular phone link, a radio frequency (RF) link, and/or other communication channels. Computer program instructions representing the block diagram and/or flowcharts herein may be loaded onto a computer, programmable data processing apparatus, or processing devices to cause a series of operations performed thereon to produce a computer implemented process. In some embodiments, processing instructions for example 300 (FIG. 3), example 600 (FIG. 6), and example 800 (FIG. 8) and processing instructions for process 1100 (FIG. 11) may be stored as program instructions on the memory 1203, storage device 1204 and the removable storage device 1205 for execution by the processor 1201.

[0053] FIG. 13 shows a block diagram for a process 1300 for context-based model selection and activation, according to some embodiments. In block 1310, objects and activities are detected in the vicinity of a user using a camera (e.g., camera 205, FIG. 2). In block 1320, based on the detection of objects and activities, process 1300 selects at least one model (e.g., a model 336, FIG. 3) from multiple models (e.g., using model management processing 212). In block 1330, also based on the detection, at least one provider (e.g., a mobile application, a web service, or a collection of information such as a database) is selected from multiple providers and the results are visually classified (e.g., using visual classifier results 211 and the activation engine 310). In block 1340, information is received from sensors (e.g., sensors 230), speech/voice input (e.g., speech/voice input 231), Internet (e.g., Internet 232) and electronic device applications (e.g., installed applications 233). In block 1350, the information received in block 1340 is classified (e.g., by context generation processing 220) into static context (e.g., static context 222), dynamic context (e.g., dynamic context 223), learned context (e.g., learned context 224) and inferred context (e.g., inferred context 225). In block 1360, the static context, dynamic context, learned context, inferred context and the visually classified results are analyzed (e.g., by the context analyzer 221). In block 1370, intent is predicted (e.g., using the intent prediction processing 240) based on the analysis of the static con-

text, dynamic context, learned context, inferred context and the visually classified results. In block 1380, process 1300 provides for receiving layers (e.g., layers 250) and the predicted intent (e.g., from the intent prediction processing 240) for layer(s) activation and for providing a recommendation (e.g., using layer activation and recommendation processing 255, the activation engine 310), and providing the activated layer(s) and recommendation to an application (e.g., camera application 260).

[0054] Embodiments have been described with reference to flowchart illustrations and/or block diagrams of methods, apparatus (systems) and computer program products. Each block of such illustrations/diagrams, or combinations thereof, can be implemented by computer program instructions. The computer program instructions when provided to a processor produce a machine, such that the instructions, which execute via the processor create means for implementing the functions/operations specified in the flowchart and/or block diagram. Each block in the flowchart /block diagrams may represent a hardware and/or software processor/process or logic. In alternative implementations, the functions noted in the blocks may occur out of the order noted in the figures, concurrently, etc.

[0055] The terms "computer program medium," "computer usable medium," "computer readable medium", and "computer program product," are used to generally refer to media such as main memory, secondary memory, removable storage drive, a hard disk installed in hard disk drive, and signals. These computer program products are means for providing software to the computer system. The computer readable medium allows the computer system to read data, instructions, messages or message packets, and other computer readable information from the computer readable medium. The computer readable medium, for example, may include non-volatile memory, such as a floppy disk, ROM, flash memory, disk drive memory, a CD-ROM, and other permanent storage. It is useful, for example, for transporting information, such as data and computer instructions, between computer systems. Computer program instructions may be stored in a computer readable medium that can direct a computer, other programmable data processing apparatus, or other devices to function in a particular manner, such that the instructions stored in the computer readable medium produce an article of manufacture including instructions which implement the function/act specified in the flowchart and/or block diagram block or blocks.

[0056] As will be appreciated by one skilled in the art, aspects of the embodiments may be embodied as a system, method or computer program product. Accordingly, aspects of the embodiments may take the form of an entirely hardware embodiment, an entirely software embodiment (including firmware, resident software, micro-code, etc.) or an embodiment combining software and hardware aspects that may all generally be referred to herein as a "circuit," "processor" or "system." Furthermore, aspects of the embodiments may take the form of

a computer program product embodied in one or more computer readable medium(s) having computer readable program code embodied thereon.

[0057] Any combination of one or more computer readable medium(s) may be utilized. The computer readable medium may be a computer readable storage medium. A computer readable storage medium may be, for example, but not limited to, an electronic, magnetic, optical, electromagnetic, infrared, or semiconductor system, apparatus, or device, or any suitable combination of the foregoing. More specific examples (a non-exhaustive list) of the computer readable storage medium would include the following: an electrical connection having one or more wires, a portable computer diskette, a hard disk, a random access memory (RAM), a read-only memory (ROM), an erasable programmable read-only memory (EPROM or Flash memory), an optical fiber, a portable compact disc read-only memory (CD-ROM), an optical storage device, a magnetic storage device, or any suitable combination of the foregoing. In the context of this document, a computer readable storage medium may be any tangible medium that can contain, or store a program for use by or in connection with an instruction execution system, apparatus, or device.

[0058] Computer program code for carrying out operations for aspects of one or more embodiments may be written in any combination of one or more programming languages, including an object-oriented programming language such as Java, Smalltalk, C++ or the like and conventional procedural programming languages, such as the "C" programming language or similar programming languages. The program code may execute entirely on the user's computer, partly on the user's computer, as a stand-alone software package, partly on the user's computer and partly on a remote computer or entirely on the remote computer or server. In the latter scenario, the remote computer may be connected to the user's computer through any type of network, including a local area network (LAN) or a wide area network (WAN), or the connection may be made to an external computer (for example, through the Internet using an Internet Service Provider).

[0059] Aspects of one or more embodiments are described above with reference to flowchart illustrations and/or block diagrams of methods, apparatus (systems) and computer program products. It will be understood that each block of the flowchart illustrations and/or block diagrams, and combinations of blocks in the flowchart illustrations and/or block diagrams, can be implemented by computer program instructions. These computer program instructions may be provided to a special purpose computer, or other programmable data processing apparatus to produce a machine, such that the instructions, which execute via the processor of the computer or other programmable data processing apparatus, create means for implementing the functions/acts specified in the flowchart and/or block diagram block or blocks.

[0060] These computer program instructions may also

be stored in a computer readable medium that can direct a computer, other programmable data processing apparatus, or other devices to function in a particular manner, such that the instructions stored in the computer readable medium produce an article of manufacture including instructions which implement the function/act specified in the flowchart and/or block diagram block or blocks.

[0061] The computer program instructions may also be loaded onto a computer, other programmable data processing apparatus, or other devices to cause a series of operational steps to be performed on the computer, other programmable apparatus or other devices to produce a computer implemented process such that the instructions which execute on the computer or other programmable apparatus provide processes for implementing the functions/acts specified in the flowchart and/or block diagram block or blocks.

[0062] The flowchart and block diagrams in the Figures illustrate the architecture, functionality, and operation of possible implementations of systems, methods, and computer program products according to various embodiments. In this regard, each block in the flowchart or block diagrams may represent a process, segment, or portion of instructions, which comprises one or more executable instructions for implementing the specified logical function(s). In some alternative implementations, the functions noted in the block may occur out of the order noted in the figures. For example, two blocks shown in succession may, in fact, be executed substantially concurrently, or the blocks may sometimes be executed in the reverse order, depending upon the functionality involved. It will also be noted that each block of the block diagrams and/or flowchart illustration, and combinations of blocks in the block diagrams and/or flowchart illustration, can be implemented by special purpose hardware-based systems that perform the specified functions or acts or carry out combinations of special purpose hardware and computer instructions.

[0063] References in the claims to an element in the singular is not intended to mean "one and only" unless explicitly so stated, but rather "one or more."

[0064] The terminology used herein is for the purpose of describing particular embodiments only and is not intended to be limiting of the invention. As used herein, the singular forms "a", "an" and "the" are intended to include the plural forms as well, unless the context clearly indicates otherwise. It will be further understood that the terms "comprises" and/or "comprising," when used in this specification, specify the presence of stated features, integers, steps, operations, elements, and/or components, but do not preclude the presence or addition of one or more other features, integers, steps, operations, elements, components, and/or groups thereof.

Claims

1. A method, comprising:

retrieving, by a device (52), contextual information based on at least one of an image, the device (52), user context, or a combination thereof; identifying two or more models from a plurality of learning models based on the contextual information and at least one object recognized in an image based on at least one visual recognition model;

activating the two or more models for the at least one object recognized in the image; updating the contextual information by adding at least one visual label identified by the two or more models to the contextual information; activating an additional model from the plurality of learning models based on the updated contextual information; and displaying information comprising at least one icon at the device (52), the at least one icon being associated with at least one of an application, a service, or a combination thereof providing additional information, wherein the additional information is associated with the additional model.

2. The method of claim 1, wherein the contextual information comprises at least one of: current time, velocity of the device (52), sensor information, user speech information, textual information, device application information, and location information, and the two or more models each comprises at least one of a machine model, a deep learning model, an object recognition model, and a generic model.
3. The method of claim 1, wherein displaying the at least one icon at the device (52) is based on at least one of a ranking of relevance of the additional model, source of the additional model, or a combination thereof, wherein the source indicates a storage within the device (52) and/or Internet (232).
4. The method of claim 1, further comprising:

activating the two or more models based on at least one of relevance of the two or more models, input from a user interface of the device (52), or a combination thereof.
5. The method of claim 1, further comprising:

determining a relevance of the two or more models based on a visual label via a computation of at least one of: semantic similarity, visual label confidence, appearance in a specified set, or a combination thereof.
6. The method of claim 2, further comprising:

inferring the contextual information based on data from an external service and sensors coupled to the device (52), wherein the contextual information further includes intent and state of a user of the device

(52).

7. The method of claim 2, further comprising:
determining whether the contextual information is related to the device (52) being in a moving vehicle, and in such a case providing less augmented reality, AR, information for objects displayed on the device (52).

8. An electronic device (52) comprising:

a memory (1203) configured to store instructions;
at least one processor (1201) configured to execute the instructions to:

retrieve contextual information based on at least one of an image, the electronic device (52), user context, or a combination thereof;
identify two or more models from a plurality of models based on the contextual information and at least one object recognized in an image based on at least one visual recognition model;

activate the two or more models for the at least one object recognized in the image;
update the contextual information by adding at least one visual label identified by the two or more models to the contextual information;

activate an additional model from the plurality of learning models based on the updated contextual information; and
display information comprising at least one icon at the electronic device (52), the at least one icon being associated with at least one of an application, a service, or a combination thereof providing additional information, wherein the additional information is associated with the additional model.

9. The electronic device (52) of claim 8, wherein:

the contextual information comprises at least one of: current time, velocity of the device (52), sensor information, user speech information, textual information, device application information, and location information;

wherein the processor (1201) is further configured to display the at least one icon at the electronic device (52) is based on at least one of a ranking of relevance of the additional model, a source of the additional model, or a combination thereof, wherein the source indicates a storage within the device (52) and/or Internet (232); and
wherein each of the two or more models comprises at least one of a machine model, a deep learning model, an object recognition model,

and a generic model.

10. The electronic device (52) of claim 9, wherein the at least one processor is further configured to execute the instructions to:

activate the two or more models based on at least one of relevance of each of the two or more models, input from a user interface of the electronic device (52), or a combination thereof; and
determine the relevance of the two or more models based on a visual label via a computation of at least one of: semantic similarity, visual label confidence, appearance in a specified set, or a combination thereof.

11. The electronic device (52) of claim 9, wherein the at least one processor is further configured to execute the instructions to:

infer the contextual information based on data from an external service and sensors coupled to the electronic device (52), wherein the contextual information further includes intent and state of a user of the electronic device (52).

12. A computer readable medium comprising program code that, when executed by at least one processor, causes the at least one processor to perform a method of any one of claims 1 to 7.

Patentansprüche

1. Verfahren, umfassend:

Abrufen, durch eine Vorrichtung (52), von kontextabhängigen Informationen auf der Grundlage mindestens eines aus einem Bild, der Vorrichtung (52), dem Benutzerkontext oder einer Kombination daraus;

Identifizieren von zwei oder mehr Modellen aus einer Mehrzahl von Lernmodellen auf der Grundlage der kontextabhängigen Informationen und mindestens eines Objekts, erkannt in einem Bild auf der Grundlage mindestens eines visuellen Erkennungsmodells;

Aktivieren der zwei oder mehr Modelle für das mindestens eine in dem Bild erkannte Objekt;
Aktualisieren der kontextabhängigen Informationen durch Hinzufügen mindestens eines visuellen Labels, identifiziert durch die zwei oder mehr Modelle, zu den kontextabhängigen Informationen;

Aktivieren eines zusätzlichen Modells aus der Mehrzahl von Lernmodellen auf der Grundlage der aktualisierten kontextabhängigen Informationen; und

Anzeigen von Informationen, umfassend min-

- destens ein Icon, auf der Vorrichtung (52), wobei das mindestens eine Icon mit mindestens einem aus einer Anwendung, einem Dienst oder einer Kombination daraus, bereitstellend zusätzliche Informationen, verknüpft ist, wobei die zusätzlichen Informationen mit dem zusätzlichen Modell verknüpft sind.
2. Verfahren nach Anspruch 1, wobei die kontextabhängigen Informationen mindestens eines aus der aktuellen Zeit, der Geschwindigkeit der Vorrichtung (52), Sensorinformationen, Benutzersprachinformationen, Textinformationen, Vorrichtungsanwendungsinformationen und Ortsinformationen umfassen und die zwei oder mehr Modelle jeweils mindestens eines aus einem Maschinenmodell, einem Modell des tiefen Lernens, einem Objekterkennungsmodell und einem generischen Modell umfassen.
 3. Verfahren nach Anspruch 1, wobei das Anzeigen des mindestens einen Icons auf der Vorrichtung (52) auf mindestens einem aus einer Rangordnung der Relevanz des zusätzlichen Modells, einer Quelle des zusätzlichen Modells oder einer Kombination daraus basiert, wobei die Quelle einen Speicher innerhalb der Vorrichtung (52) und/oder des Internets (232) angibt.
 4. Verfahren nach Anspruch 1, ferner umfassend: Aktivieren der zwei oder mehr Modelle auf der Grundlage mindestens eines aus der Relevanz der zwei oder mehr Modelle, der Eingabe von einer Benutzeroberfläche der Vorrichtung (52) oder einer Kombination daraus.
 5. Verfahren nach Anspruch 1, ferner umfassend: Bestimmen einer Relevanz der zwei oder mehr Modelle auf der Grundlage eines visuellen Labels mittels einer Berechnung mindestens eines aus der semantischen Nähe, des Vertrauens des visuellen Labels, des Erscheinens in einer spezifizierten Menge oder einer Kombination daraus.
 6. Verfahren nach Anspruch 2, ferner umfassend: Inferieren der kontextabhängigen Informationen auf der Grundlage von Daten aus einem externen Dienst und an die Vorrichtung (52) gekoppelten Sensoren, wobei die kontextabhängigen Informationen ferner die Absicht und den Zustand eines Benutzers der Vorrichtung (52) aufweisen.
 7. Verfahren nach Anspruch 2, ferner umfassend: Bestimmen, ob die kontextabhängigen Informationen mit dem Vorhandensein der Vorrichtung (52) in einem sich bewegenden Fahrzeug in Beziehung stehen, und in einem solchen Fall, Bereitstellen von weniger Informationen der erweiterten Realität, *augmented reality*- AR, für auf der Vorrichtung (52) angezeigte Objekte.
 8. Elektronische Vorrichtung (52), umfassend:
 - einen Speicher (1203), der dazu konfiguriert ist, Befehle zu speichern;
 - mindestens einen Prozessor (1201), der dazu konfiguriert ist, die Befehle auszuführen, um:
 - kontextabhängige Informationen auf der Grundlage mindestens eines aus einem Bild, der elektronischen Vorrichtung (52), dem Benutzerkontext oder einer Kombination daraus abzurufen;
 - zwei oder mehr Modelle aus einer Mehrzahl von Modellen auf der Grundlage der kontextabhängigen Informationen und mindestens eines Objekts, erkannt in einem Bild auf der Grundlage mindestens eines visuellen Erkennungsmodells, zu identifizieren;
 - die zwei oder mehr Modelle für das mindestens eine in dem Bild erkannte Objekt zu aktivieren;
 - die kontextabhängigen Informationen durch Hinzufügen mindestens eines visuellen Labels, identifiziert durch die zwei oder mehr Modelle, zu den kontextabhängigen Informationen zu aktualisieren;
 - ein zusätzliches Modell aus der Mehrzahl von Lernmodellen auf der Grundlage der aktualisierten kontextabhängigen Informationen zu aktivieren; und
 - Informationen, umfassend mindestens ein Icon, auf der elektronischen Vorrichtung (52) anzuzeigen, wobei das mindestens eine Icon mit mindestens einem aus einer Anwendung, einem Dienst oder einer Kombination daraus, bereitstellend zusätzliche Informationen, verknüpft ist, wobei die zusätzlichen Informationen mit dem zusätzlichen Modell verknüpft sind.
 9. Elektronische Vorrichtung (52) nach Anspruch 8, wobei:
 - die kontextabhängigen Informationen mindestens eines aus der aktuellen Zeit, der Geschwindigkeit der Vorrichtung (52), Sensorinformationen, Benutzersprachinformationen, Textinformationen, Vorrichtungsanwendungsinformationen und Ortsinformationen umfassen;
 - wobei der Prozessor (1201) ferner dazu konfiguriert ist, das mindestens eine Icon auf der elektronischen Vorrichtung (52) auf der Grundlage mindestens eines aus einer Rangordnung der Relevanz des zusätzlichen Modells, einer Quelle des zusätzlichen Modells oder einer Kombination daraus anzuzeigen, wobei die

- Quelle einen Speicher innerhalb der Vorrichtung (52) und/oder des Internets (232) angibt; und wobei jedes der zwei oder mehr Modelle mindestens eines aus einem Maschinenmodell, einem Modell des tiefen Lernens, einem Objekterkennungsmodell und einem generischen Modell umfasst.
10. Elektronische Vorrichtung (52) nach Anspruch 9, wobei der mindestens eine Prozessor ferner dazu konfiguriert ist, die Befehle auszuführen, um:
- die zwei oder mehr Modelle auf der Grundlage mindestens eines aus der Relevanz jedes der zwei oder mehr Modelle, der Eingabe von einer Benutzeroberfläche der elektronischen Vorrichtung (52) oder einer Kombination daraus zu aktivieren; und
- die Relevanz der zwei oder mehr Modelle auf der Grundlage eines visuellen Labels mittels einer Berechnung mindestens eines aus der semantischen Nähe, des Vertrauens des visuellen Labels, des Erscheinens in einer spezifizierten Menge oder einer Kombination daraus zu bestimmen.
11. Elektronische Vorrichtung (52) nach Anspruch 9, wobei der mindestens eine Prozessor ferner dazu konfiguriert ist, die Befehle auszuführen, um:
- die kontextabhängigen Informationen auf der Grundlage von Daten aus einem externen Dienst und an die elektronische Vorrichtung (52) gekoppelten Sensoren zu inferieren, wobei die kontextabhängigen Informationen ferner die Absicht und den Zustand eines Benutzers der elektronischen Vorrichtung (52) aufweisen.
12. Maschinell lesbarer Datenträger, umfassend einen Programmcode, der, wenn er durch mindestens einen Prozessor ausgeführt wird, bewirkt, dass der mindestens eine Prozessor ein Verfahren nach einem der Ansprüche 1 bis 7 durchführt.
- Revendications**
1. Procédé, comprenant :
- la récupération, par un dispositif (52), d'informations contextuelles sur la base d'au moins une image,
- le dispositif (52), le contexte utilisateur ou une combinaison de tout cela ;
- l'identification de deux modèles ou plus parmi une pluralité de modèles d'apprentissage sur la base des informations contextuelles et au moins un objet reconnu dans une image sur la base d'au moins un modèle de reconnaissance
- visuelle ;
- l'activation des deux modèles ou plus pour au moins un objet reconnu dans l'image ;
- la mise à jour des informations contextuelles en y ajoutant au moins une étiquette visuelle identifiée par les deux modèles ou plus ;
- l'activation d'un modèle supplémentaire parmi la pluralité de modèles d'apprentissage sur la base des informations contextuelles mises à jour; et
- l'affichage d'informations comprenant au moins une icône au niveau du dispositif (52), cette icône étant associée à au moins une application, un service ou une combinaison de tout cela fournissant des informations supplémentaires, où les informations supplémentaires sont associées au modèle supplémentaire.
2. Procédé selon la Revendication 1, où les informations contextuelles comprennent au moins :
- l'heure actuelle, la vitesse du dispositif (52), les informations du capteur, les informations vocales de l'utilisateur, les informations textuelles, les informations sur l'application du dispositif et les informations de localisation, et les deux modèles ou plus comprennent chacun au moins un modèle de machine, un modèle d'apprentissage profond, un modèle de reconnaissance d'objet et un modèle générique.
3. Procédé selon la Revendication 1, où l'affichage d'au moins une icône au niveau du dispositif (52) est basé sur au moins un classement de pertinence du modèle supplémentaire,
- une source du modèle supplémentaire, ou une combinaison de tout cela, où la source indique un stockage dans le dispositif (52) et/ou l'Internet (232).
4. Procédé selon la Revendication 1, comprenant en outre :
- l'activation des deux modèles ou plus sur la base d'au moins une pertinence des deux modèles ou plus,
- l'entrée à partir d'une interface utilisateur du dispositif (52), ou une combinaison de tout cela.
5. Procédé selon la Revendication 1, comprenant en outre :
- la détermination de la pertinence des deux modèles ou plus sur la base d'une étiquette visuelle via un calcul d'au moins un des éléments suivants : similarité sémantique, confiance dans l'étiquette visuelle, apparence dans un ensemble spécifié, ou une combinaison de tout cela.
6. Procédé selon la Revendication 2, comprenant en outre :
- l'inférence des informations contextuelles sur la ba-

se de données provenant d'un service externe et de capteurs couplés au dispositif (52), où les informations contextuelles comprennent en outre l'intention et l'état d'un utilisateur du dispositif (52).

7. Procédé selon la Revendication 2, comprenant en outre :

le fait de déterminer si les informations contextuelles sont liées au fait que le dispositif (52) se trouve dans un véhicule en mouvement, et dans ce cas

le fait de fournir moins d'informations en réalité augmentée, *augmented reality*, AR, pour les objets affichés sur le dispositif (52).

8. Dispositif électronique (52) comprenant :

une mémoire (1203) configurée pour stocker des instructions ;

au moins un processeur (1201) configuré pour exécuter les instructions pour :

recupérer des informations contextuelles sur la base d'au moins une image, le dispositif électronique (52), le contexte utilisateur ou une combinaison de tout cela ;

identifier deux modèles ou plus parmi une pluralité de modèles sur la base des informations contextuelles et au moins un objet reconnu dans une image sur la base d'au moins un modèle de reconnaissance visuelle ;

activer les deux modèles ou plus pour au moins un objet reconnu dans l'image ;

mettre à jour les informations contextuelles en y ajoutant au moins une étiquette visuelle identifiée par les deux modèles ou plus ;

activer un modèle supplémentaire parmi la pluralité de modèles d'apprentissage sur la base des informations contextuelles mises à jour ; et

afficher des informations comprenant au moins une icône au niveau du dispositif électronique (52), cette icône étant associée à au moins une application, un service ou une combinaison de tout cela fournissant des informations supplémentaires, où les informations supplémentaires sont associées au modèle supplémentaire.

9. Dispositif électronique (52) selon la Revendication 8, où :

les informations contextuelles comprennent au moins : l'heure actuelle, la vitesse du dispositif (52), les informations du capteur, les informations vocales de l'utilisateur, les informations

textuelles, les informations sur l'application du dispositif et les informations de localisation ;

où le processeur (1201) est en outre configuré pour afficher au moins une icône au niveau du dispositif électronique (52) et est basé sur au moins un classement de pertinence du modèle supplémentaire,

une source du modèle supplémentaire, ou une combinaison de tout cela,

où la source indique un stockage dans le dispositif (52) et/ou l'Internet (232) ; et

où chacun des deux modèles ou plus comprend au moins un modèle de machine, un modèle d'apprentissage profond, un modèle de reconnaissance d'objet et un modèle générique.

10. Dispositif électronique (52) selon la Revendication 9, où au moins un processeur est en outre configuré pour exécuter les instructions pour :

activer les deux modèles ou plus sur la base d'au moins une pertinence de chacun des deux modèles ou plus, l'entrée à partir d'une interface utilisateur du dispositif électronique (52), ou une combinaison de tout cela ; et

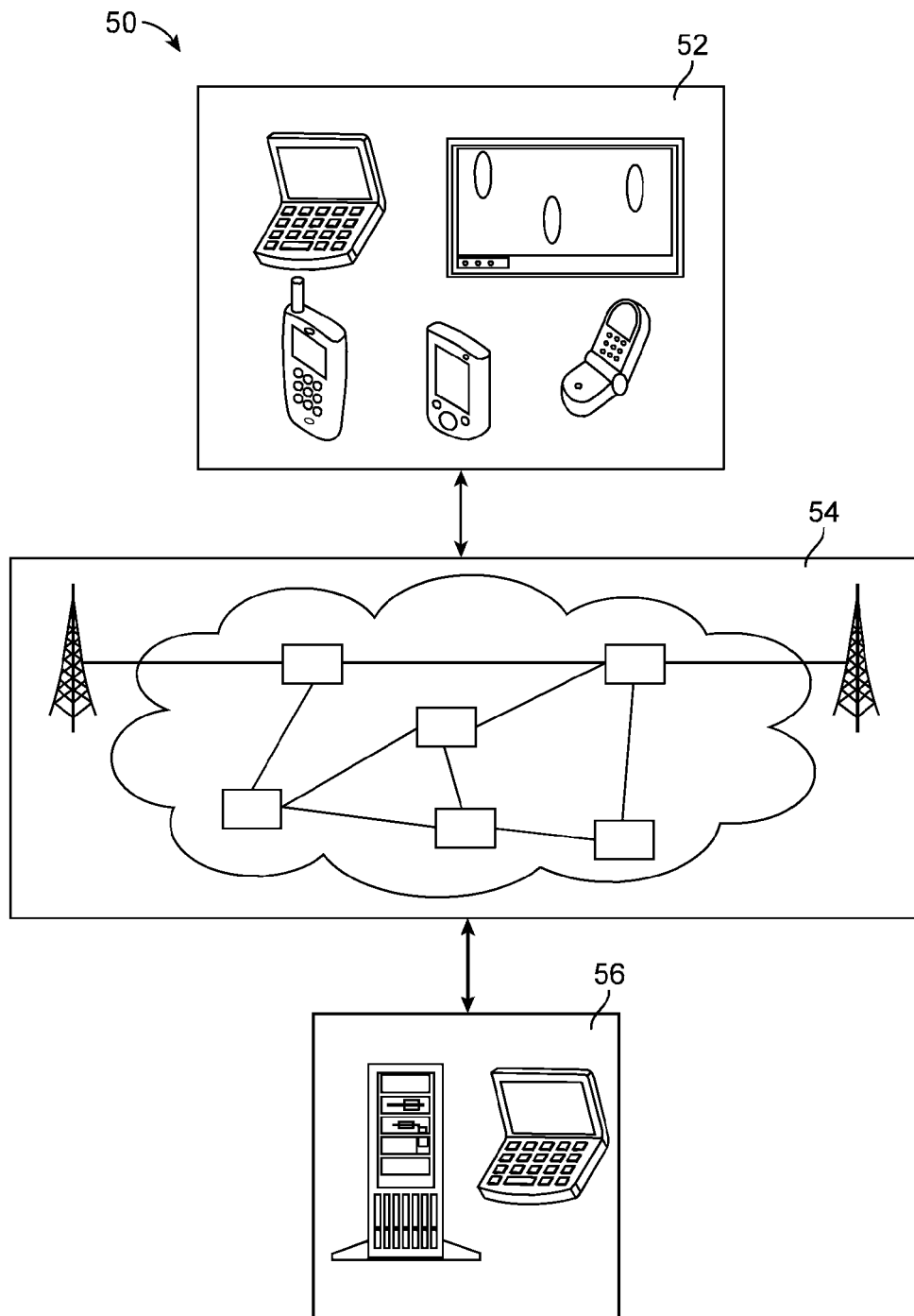
déterminer la pertinence des deux modèles ou plus sur la base d'une étiquette visuelle via un calcul d'au moins un des éléments suivants : similarité sémantique, confiance dans l'étiquette visuelle, apparence dans un ensemble spécifié, ou une combinaison de tout cela.

11. Dispositif électronique (52) selon la Revendication 9, où au moins un processeur est en outre configuré pour exécuter les instructions pour :

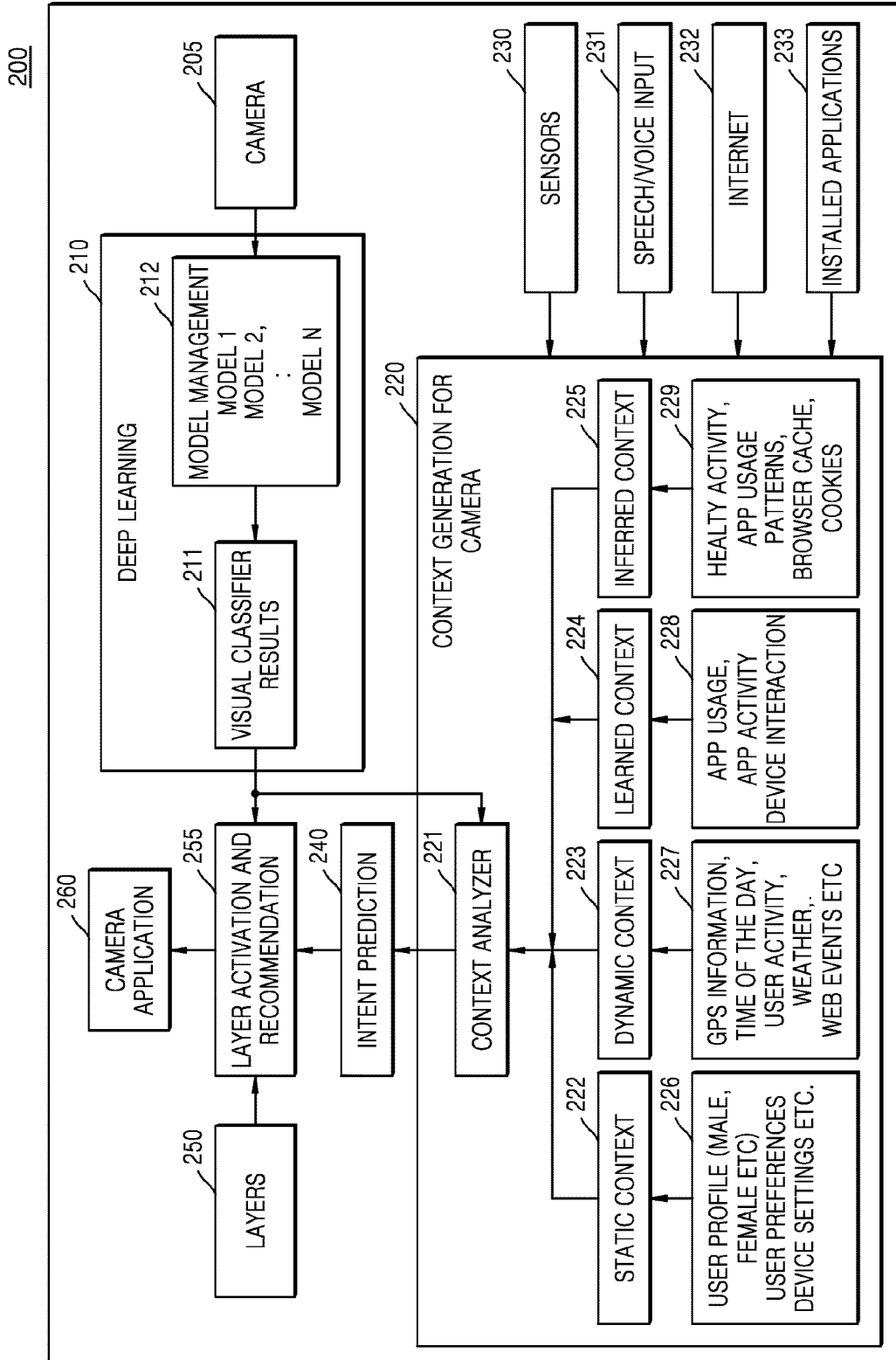
inférer les informations contextuelles sur la base de données provenant d'un service externe et de capteurs couplés au dispositif électronique (52), où les informations contextuelles comprennent en outre l'intention et l'état d'un utilisateur du dispositif électronique (52).

12. Support lisible par ordinateur comprenant un code de programme qui, lorsqu'il est exécuté par au moins un processeur, fait en sorte que ce processeur exécute un procédé selon l'une quelconque des Revendications 1 à 7.

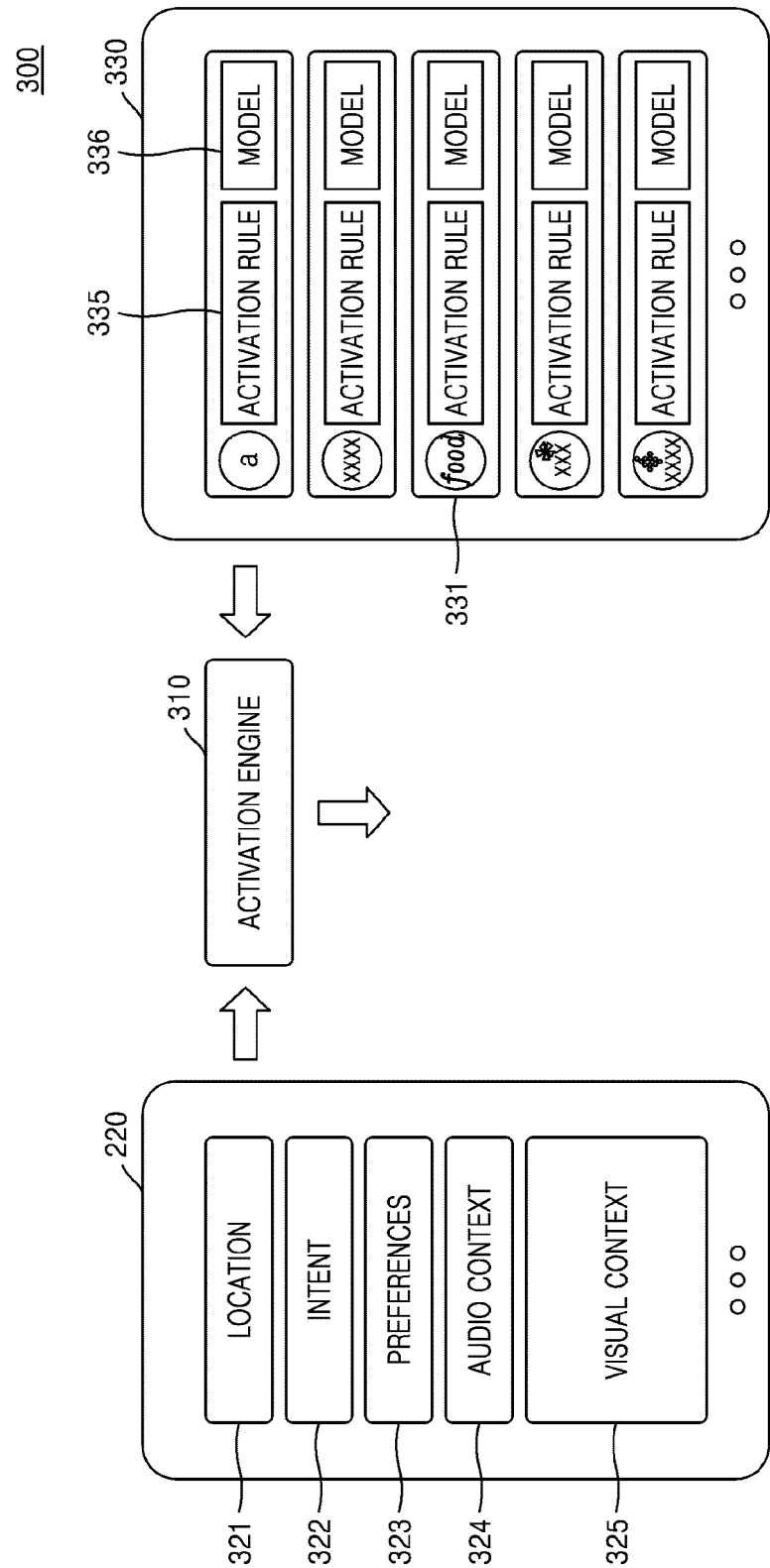
[Fig. 1]



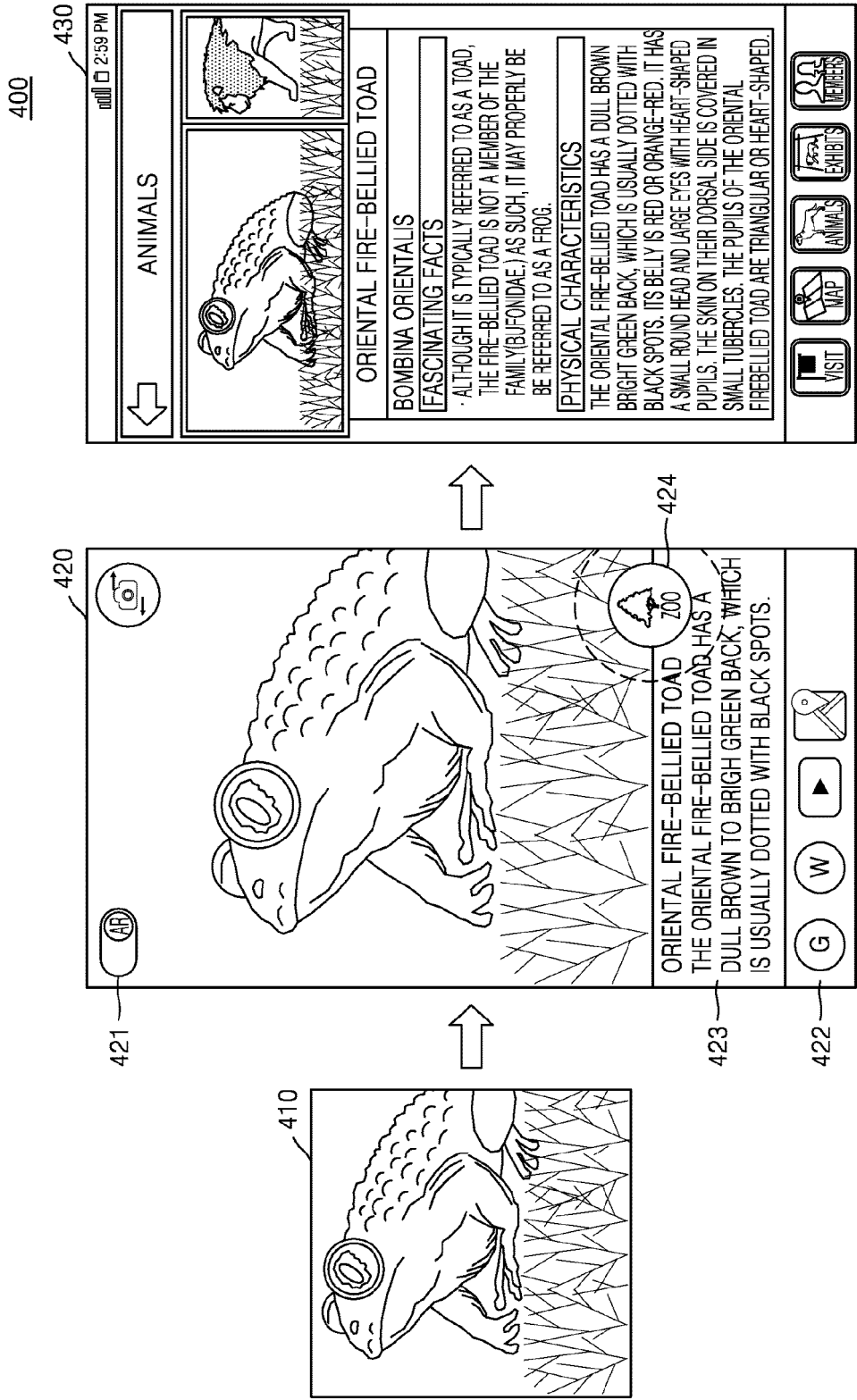
[Fig. 2]



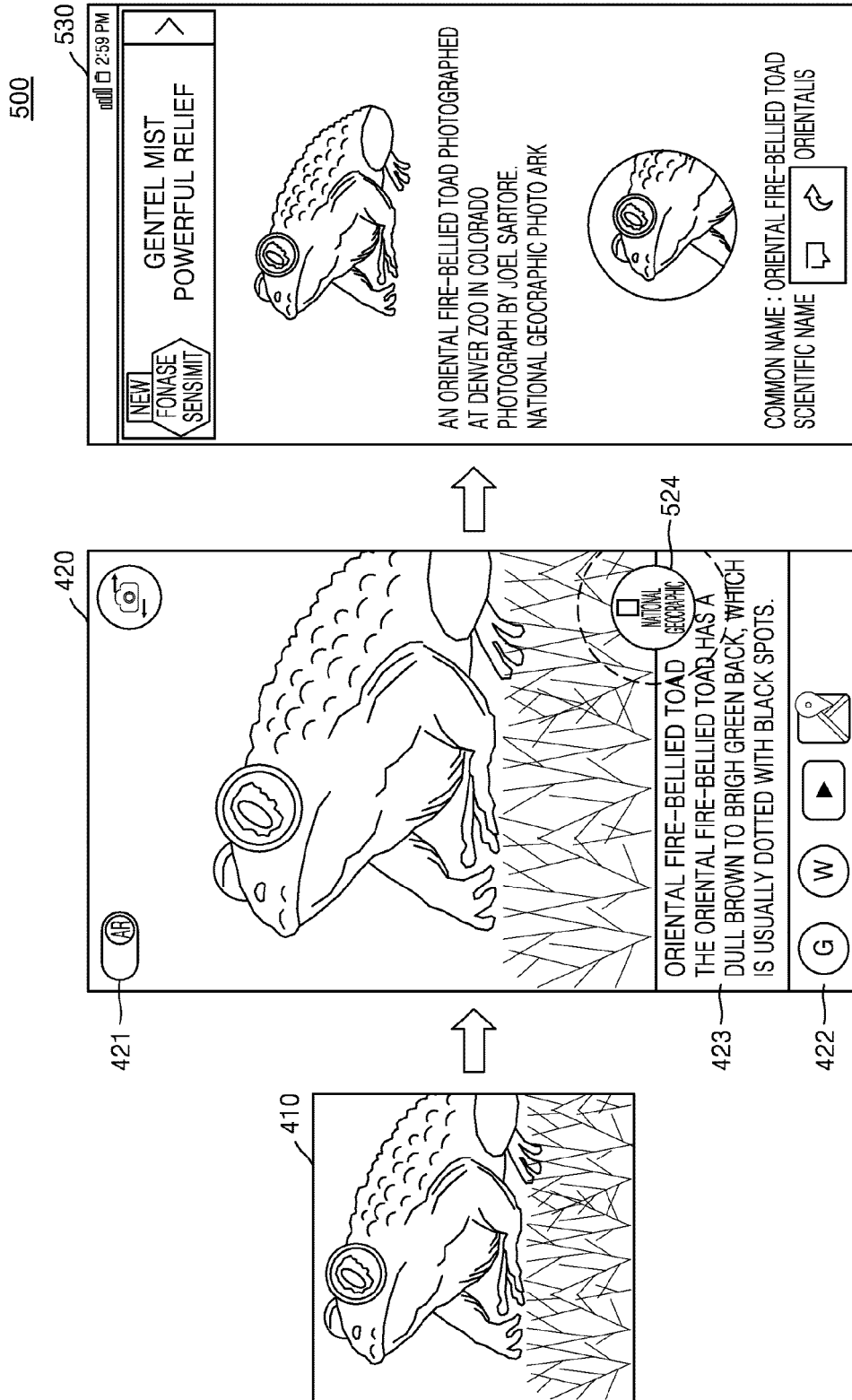
[Fig. 3]



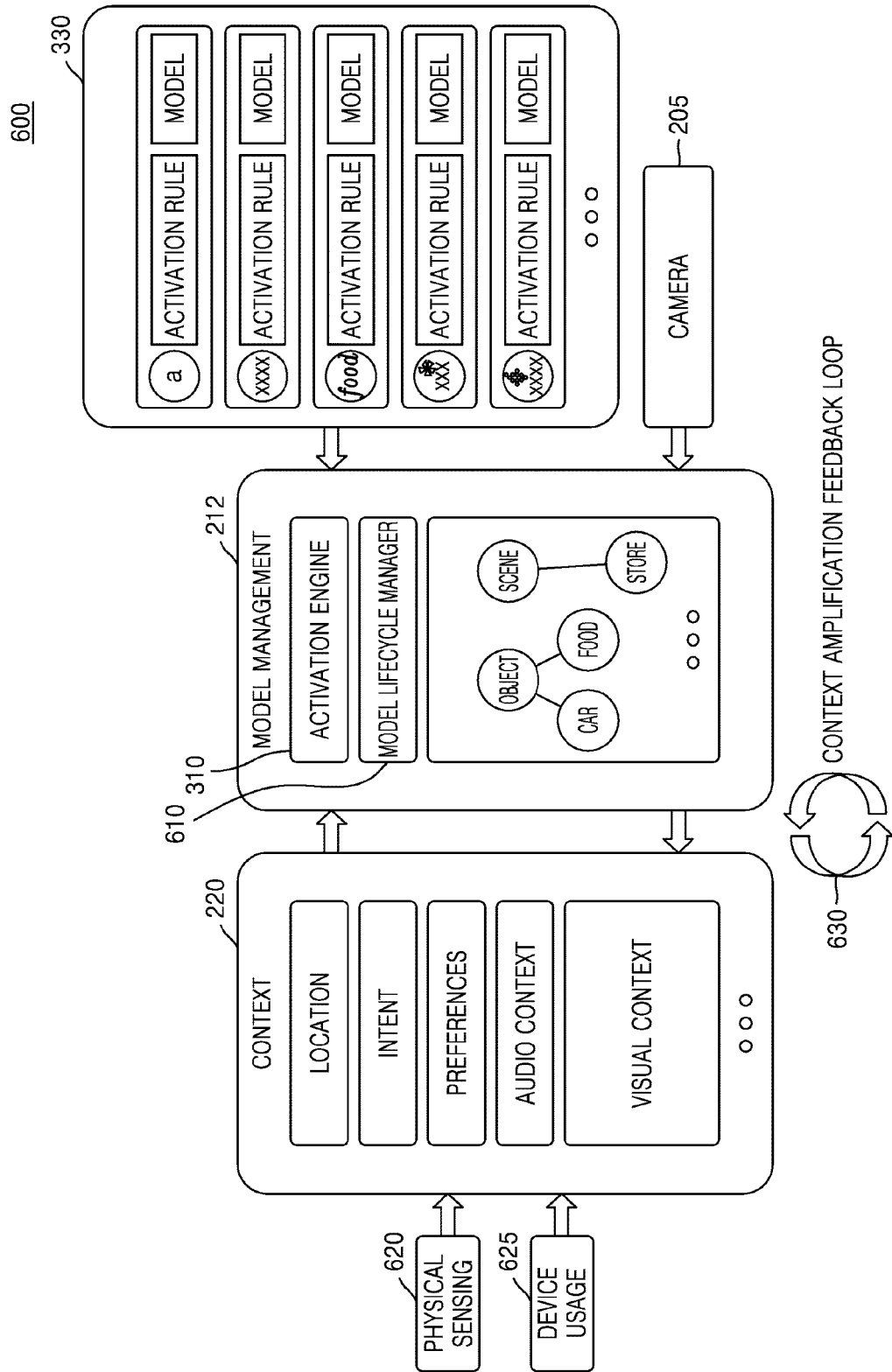
[Fig. 4]



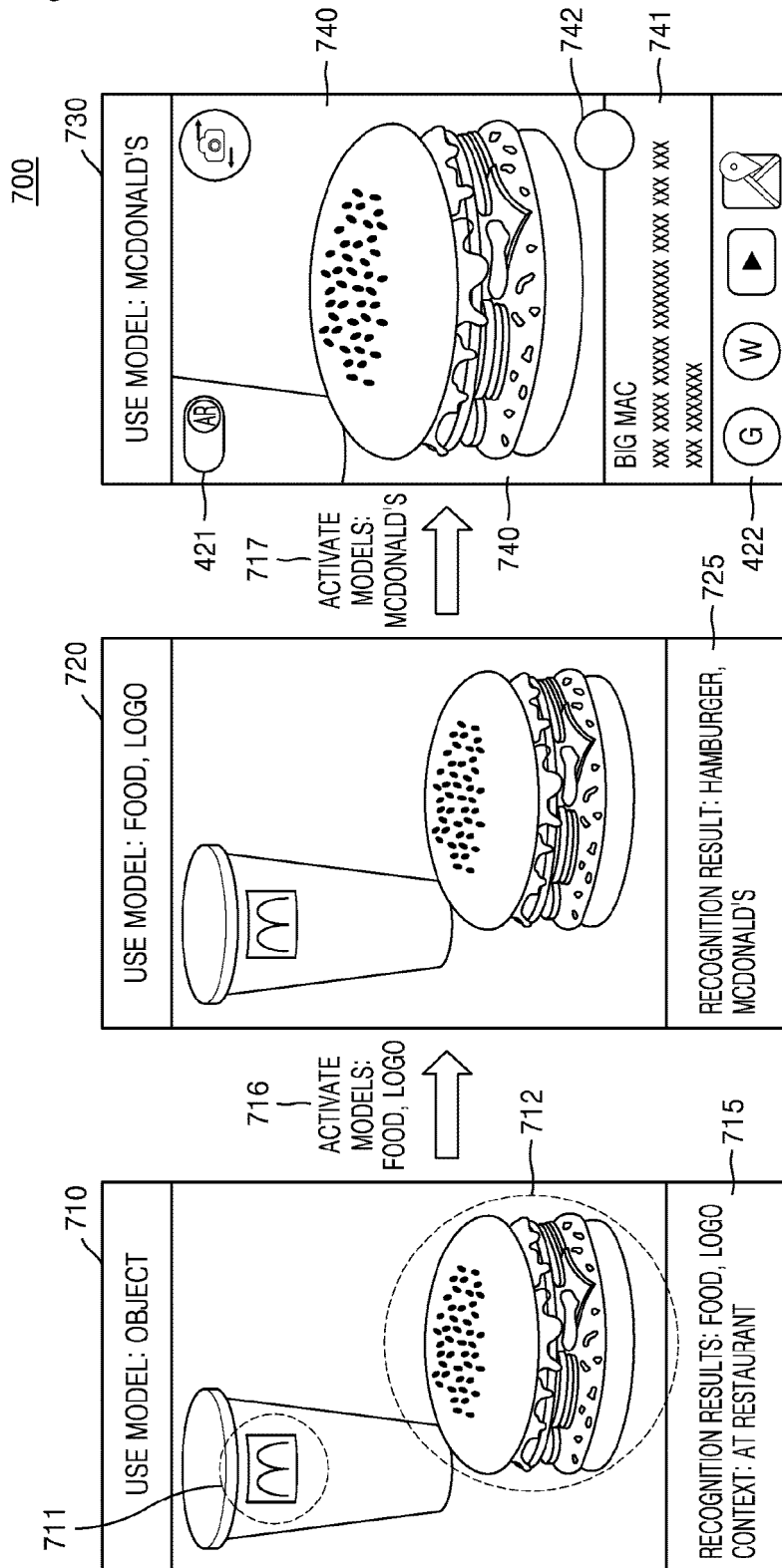
[Fig. 5]



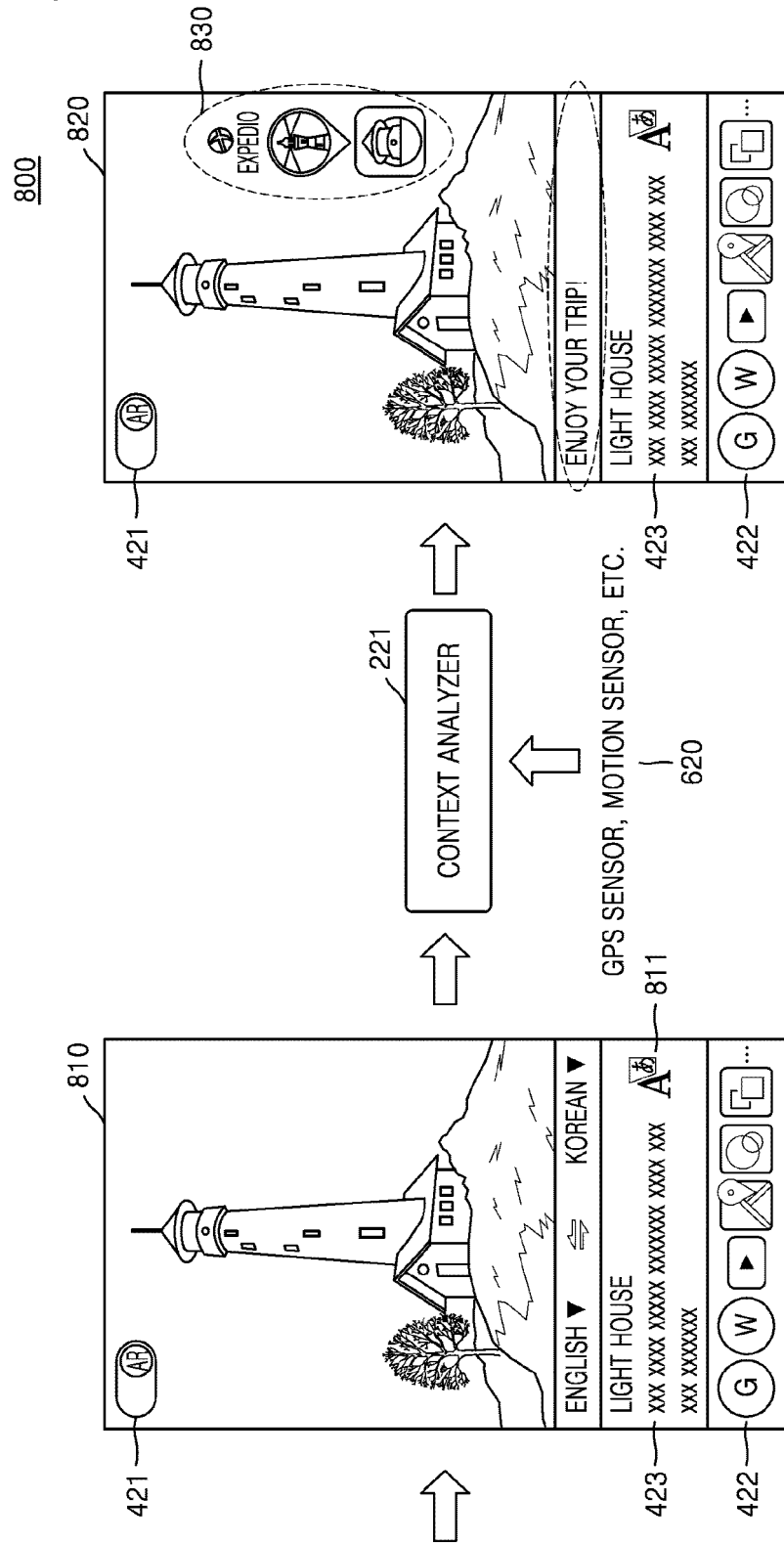
[Fig. 6]



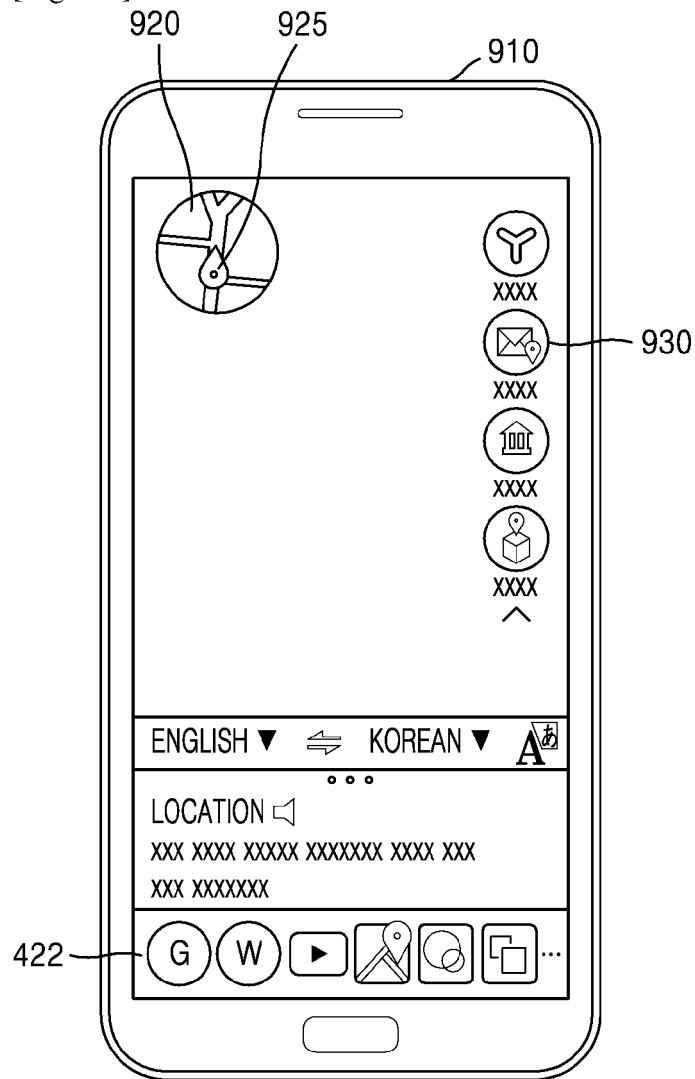
[Fig. 7]



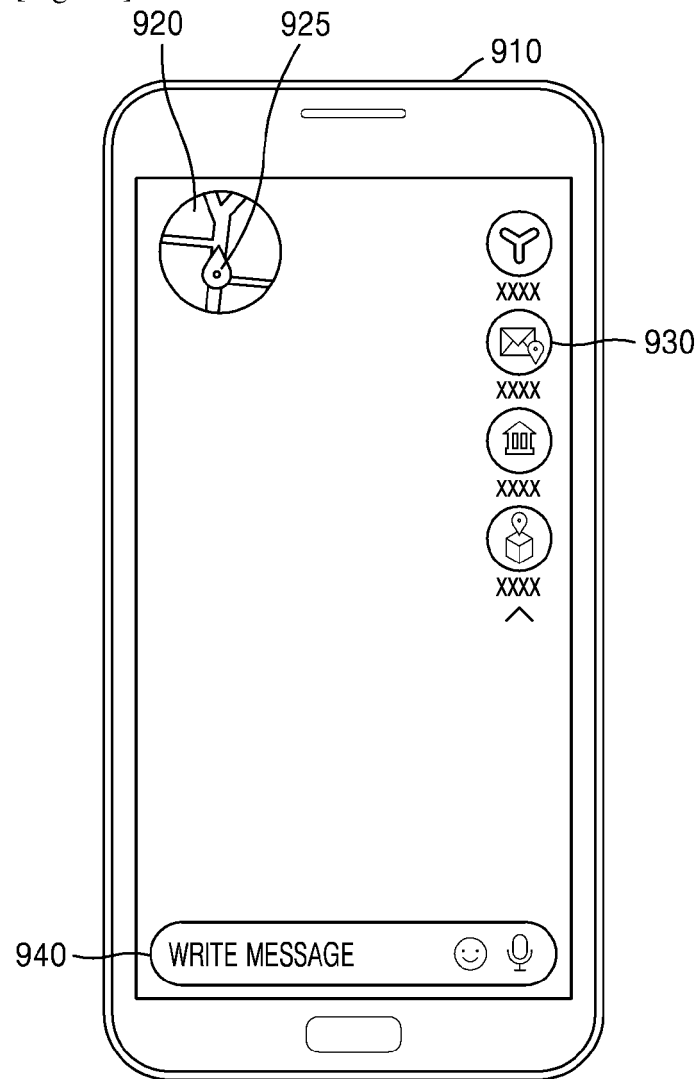
[Fig. 8]



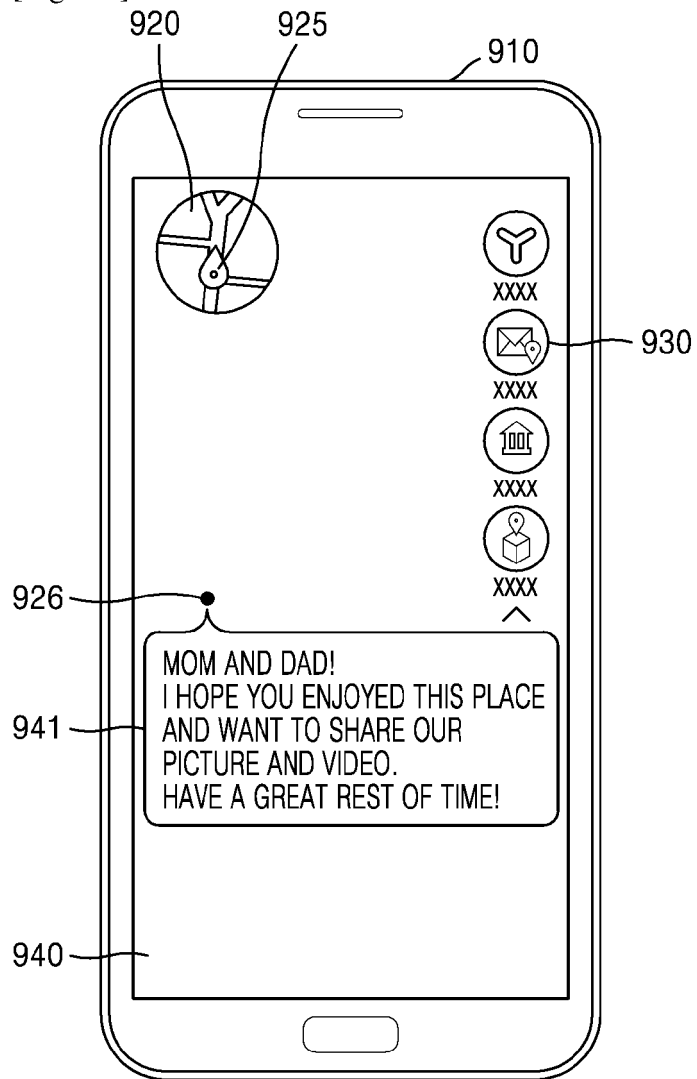
[Fig. 9A]



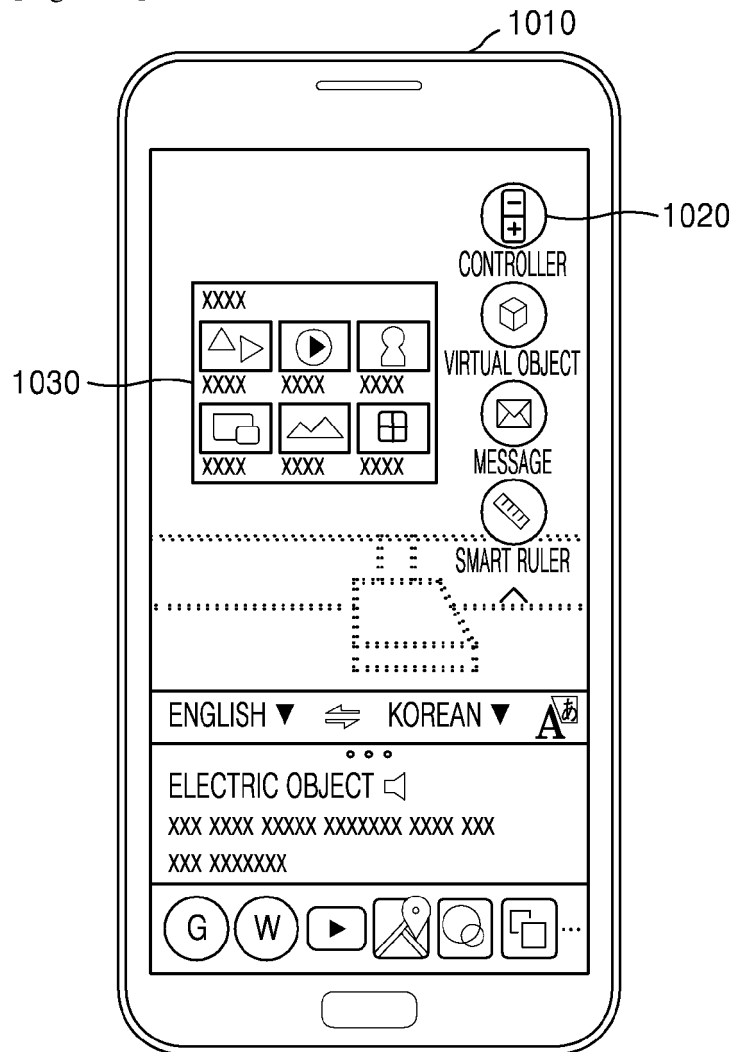
[Fig. 9B]



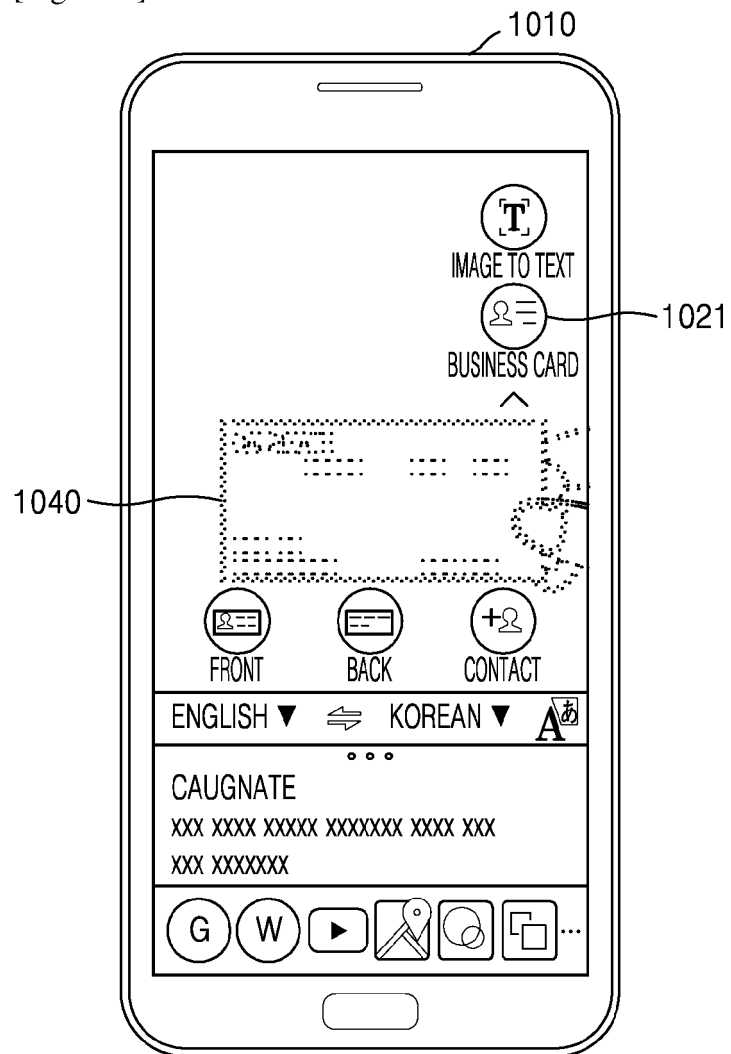
[Fig. 9C]



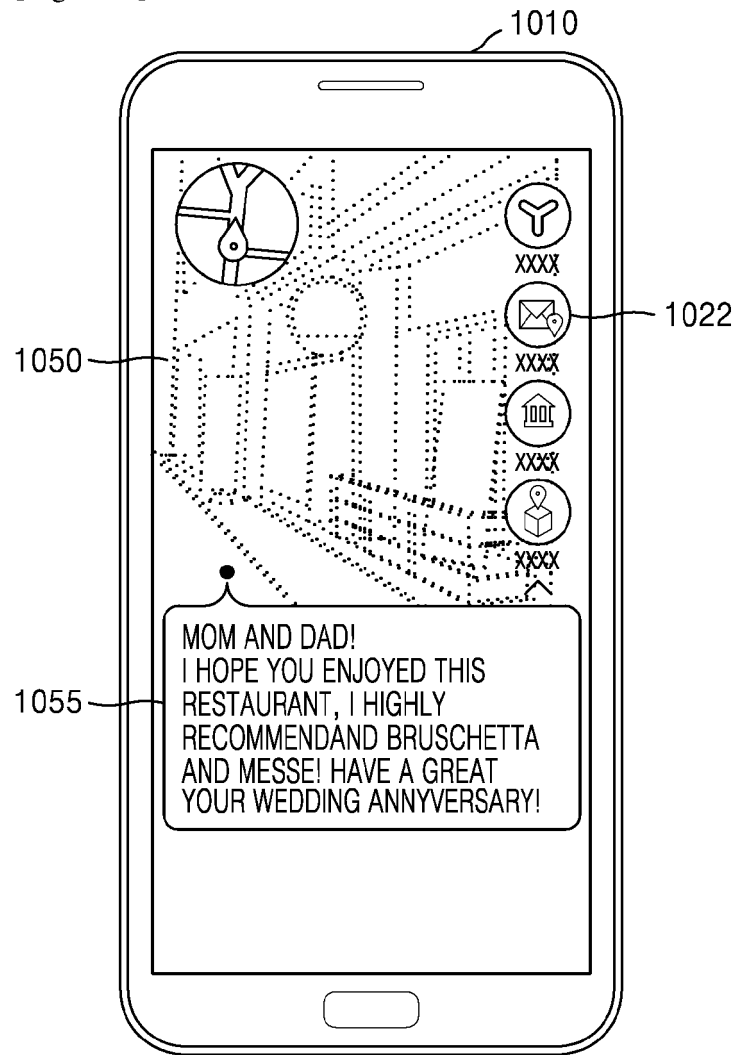
[Fig. 10A]



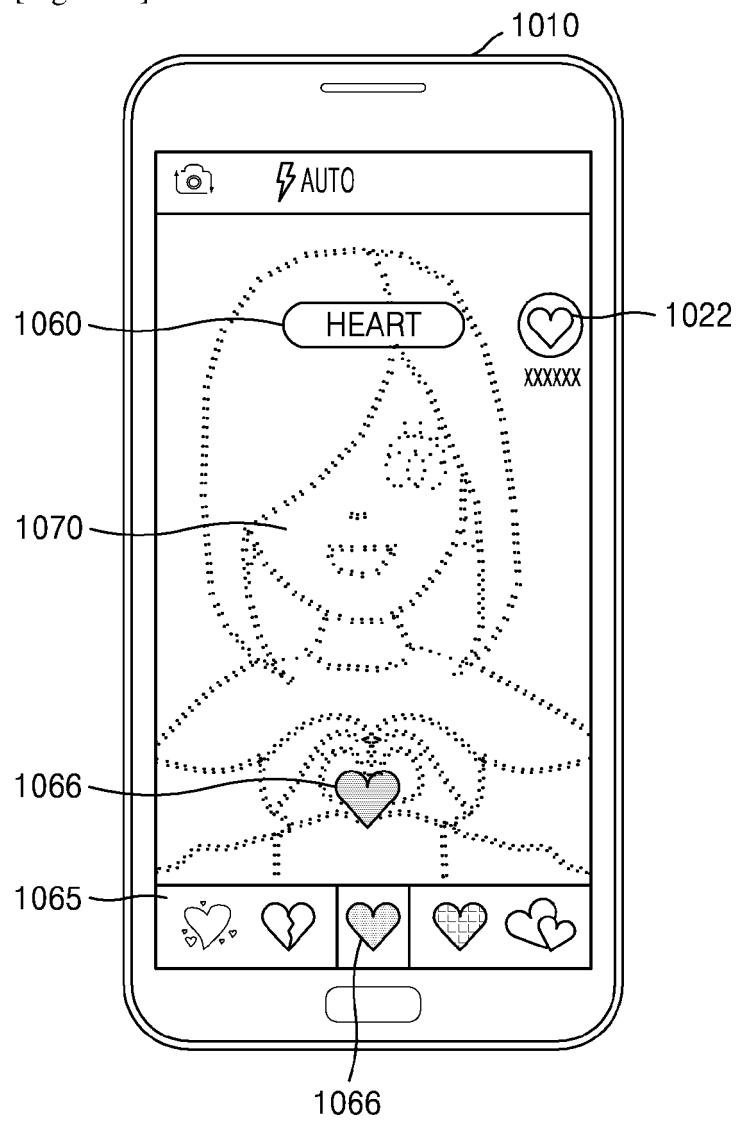
[Fig. 10B]



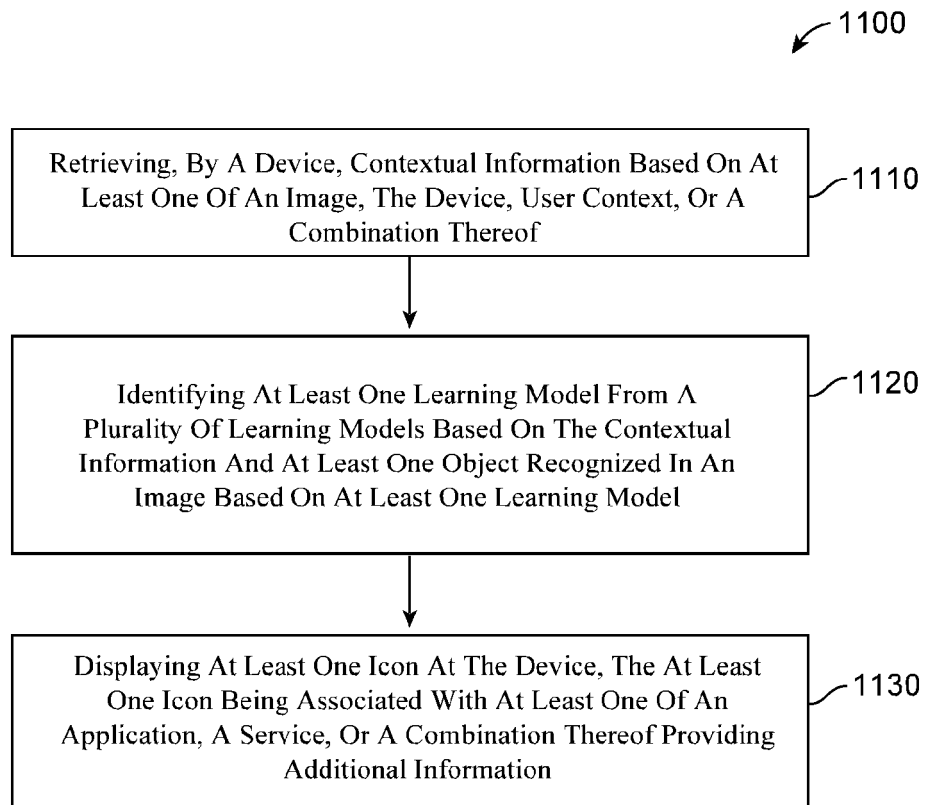
[Fig. 10C]



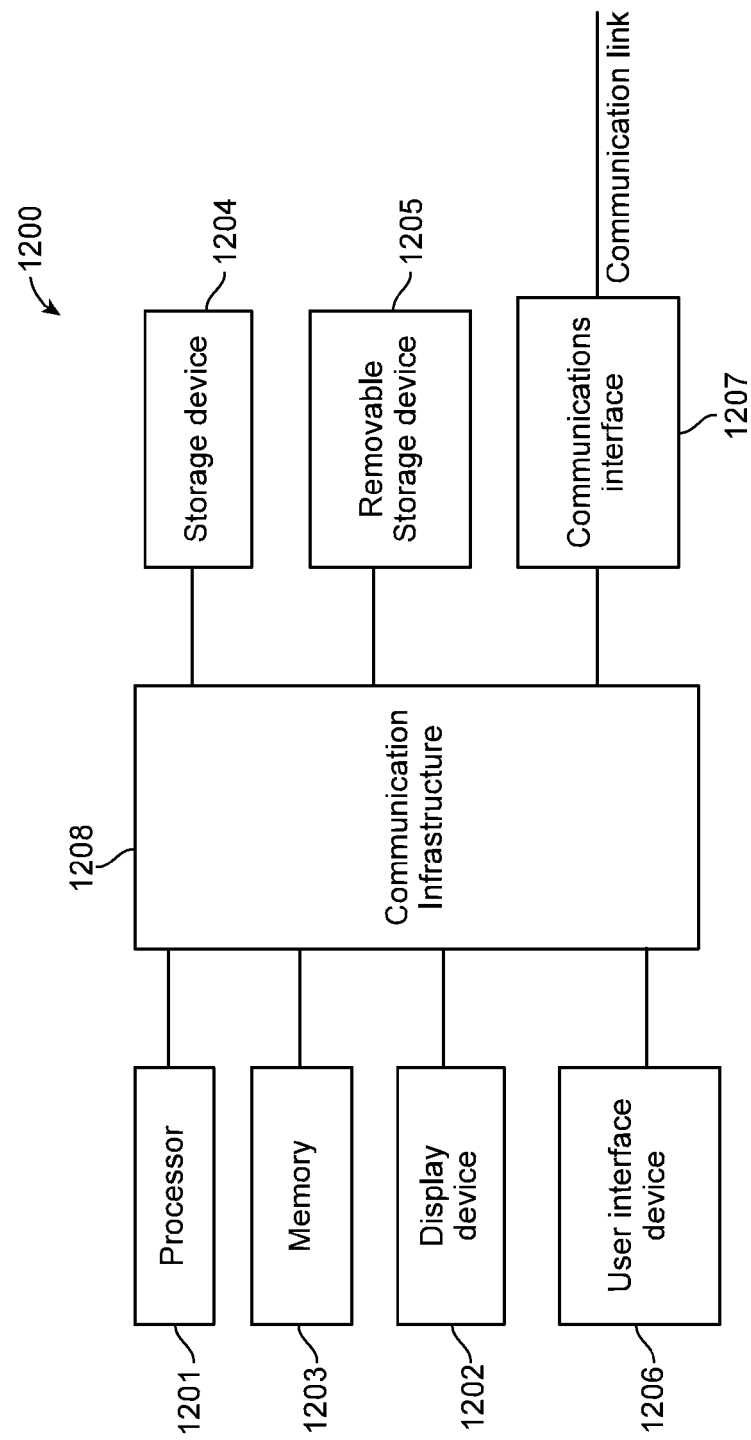
[Fig. 10D]



[Fig. 11]



[Fig. 12]



[Fig. 13]

